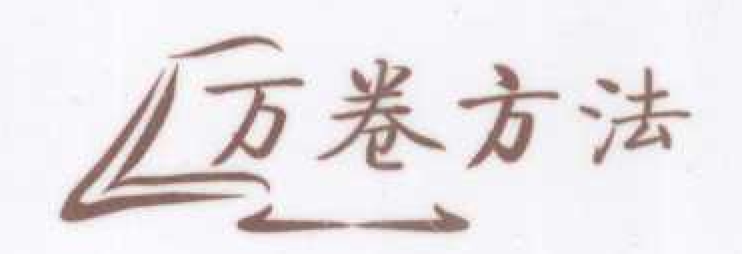


实验设计原理：
社会科学理论验证的一种路径

BUILDING EXPERIMENTS:
TESTING SOCIAL THEORY

戴维·威勒（David Willer）
亨利·沃克（Henry A. Walker） 著
杜伟宇 孟 琦 译



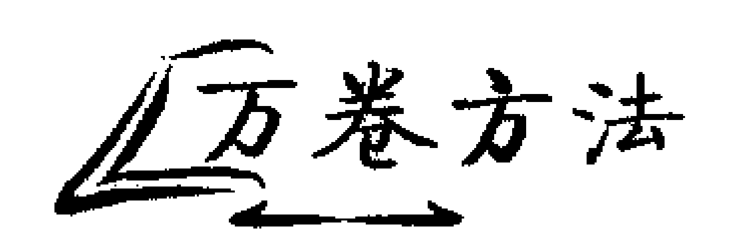
SHIYAN SHEJI YUANLI SHEHUI KEXUE LILUN YANZHENG DE YIZHONG LUJING

■ 本书介绍了实验方法在社会科学研究领域中的应用，把实验方法区分为经验驱动型和理论驱动型。前者是帮助研究者从实验现象中归纳出规律，后者是对抽象理论的实证验证。这本书对多个典型的实验进行解析，如伽利略的比萨斜塔重物下落实验、阿基米德的杠杆实验、阿施的从众实验、米尔格拉姆的权威（电击）实验等，并将自然科学中的实验与社会科学中的实验进行对比，除了总结出优秀实验设计共有的5条原则外，还以严谨的逻辑分析指出，以实验来检验理论，作为一种基本方法，完全可以像在自然科学中那样，在社会科学的研究中得到广泛的应用，并且能够对理论的发展起到巨大的推动作用。

■ 本书在不长的篇幅中却含有十分丰富的信息量，特别是从实验构建的角度讨论对一些经典理论的检验，给人许多启发。

发表及参阅相关讨论，请登录：
万卷方法与学术规范博客圈（http//q.blog.sina.com.cn/fafang）


ISBN 978-7-5624-5187-7
9 787562 451877 >
定价：20.00元



实验设计原理：社会科学理论验证的一种路径

BUILDING EXPERIMENTS: TESTING SOCIAL THEORY

戴维·威勒（David Willer）
亨利·沃克（Henry A. Walker）著
杜伟宇 孟琦 译

重庆大学出版社

版贸核渝字 2008 第(031)号

图书在版编目(CIP)数据

实验设计原理:社会科学理论验证的一种路径/(美)威勒(Willer,D.),(美)沃克(Walker,H. A.)著;杜伟宇,孟琦译. —重庆:重庆大学出版社,2010. 2
(万卷方法. 社会科学研究方法经典译丛)
书名原文:Building Experiments:Testing Social Theory
ISBN 978-7-5624-5187-7

Ⅰ. 实… Ⅱ. ①威…②沃…③杜… Ⅲ. 社会科学—科学实验—方法 Ⅳ. C33

中国版本图书馆 CIP 数据核字(2009)第 229500 号

实验设计原理:
社会科学理论验证的一种路径

戴维·威勒 David Willer
亨利·沃克 Henry A. Walker 著
杜伟宇 孟 琦 译
责任编辑:林佳木 吴文静 版式设计:林佳木
责任校对:张洪梅 责任印制:赵 晟
*
重庆大学出版社出版发行
出版人:张鸽盛
社址:重庆市沙坪坝正街 174 号重庆大学(A 区)内
邮编:400030
电话:(023) 65102378 65105781
传真:(023) 65103686 65105565
网址:http://www. cqup. com. cn
邮箱:fxk@ cqup. com. cn (营销中心)
全国新华书店经销
重庆华林天美印务有限公司印刷
*
开本:940×1360 1/32 印张:5. 125 字数:141 千
2010 年 2 月第 1 版 2010 年 2 月第 1 次印刷
印数:1—4 000
ISBN 978-7-5624-5187-7 定价:20. 00 元

译者序

实验研究跟人类的好奇心一样古老，它是科学的基石，没有实验研究就不存在真正的科学。在自然现象的研究中，实验作为一种经验认识方法，是自然科学最直接、最重要的认识基础，自然科学的每一理论的发现或检验无不借助于科学实验。在16及17世纪科学革命的时代，实验研究方法得到了极大的推动和刺激。20世纪，古典的实验研究方法一直广泛地为科学界所接纳。实验研究最早使用在物理学领域中，后来在自然科学领域广泛采用，而后以生理学的实验研究为一个发展中介，并以心理学实验为开端开始向社会科学领域逐渐推广。

在社会科学研究中，社会现象复杂多变，但实验研究所包含的实证精神却是社会研究的精髓。实验的特征、结构、程序、设计原则、信度和效度，以及实验与理论之间的关系等，都是实验方法论贡献给自然科学和社会科学的主要成果。实验研究成为典范的原因就在于它的逻辑结构。实验研究是一种很有效的社会科学研究方法，在各种类型的研究中，它是一种能最好地在变量间建立因果联系的方法。一般来说相关研究不能确定两个变量 A、B 之间是否为因果关系，这是因为有相关关系的两个变量之间可能有三种解释：第一种，A 是 B 的原因或一部分原因；第二种，B 是 A 的原因或一部分原因；第三种，A 和 B 是第三个变量 C 的原因（或结果）或一部分原因（或结果）。因此，要确定两个变量之间是因果关系，应当设计恰当的实验来寻找验证。

自然科学的研究对象是自然现象，可以在实验室创造出完

全封闭的人工环境,并在严格的控制条件下获取数据,从而发现或验证科学理论,因而较少涉及伦理、道德、法律等问题。而社会实验的研究对象是人,实验室无法创造出脱离社会结构、社会制度和社会文化的封闭环境,实验刺激必须以不伤害个人的身心健康、社会秩序和伦理道德为限度。例如,研究拥挤程度与侵犯行为之间的关系,就需要把很多人禁闭在一个狭小的环境中,但这样的禁闭不仅是违法的,而且把无辜的实验对象置于危险中也是不道德的。任何一个社会科学家都不会进行这样的实验。怎么做才能把实验方法融入到社会科学研究中去呢?

在社会问题研究的历程中,也不乏采用实验方法的知名研究。20 世纪 30 年代哈佛大学教授梅奥在美国芝加哥西部电器公司,做了一系列广为人知的实验。其中,在霍桑工厂为了提高生产效率,进行了一项旨在研究工人的生产效率是否与工作环境相关的社会实验。在实验中,根据照明等工作环境与效率相关的假设,把装配电器的工人分成两个小组:实验组和对照组,并分别安排在两个房间里工作。对照组的照明条件保持不变,实验组的照明条件被改变。还有米尔格拉姆在 20 世纪 70 年代关于服从权威的一系列实验。其中最著名的实验是,两个获得酬劳的志愿者,一个是被试,另一人是实验者的同伴。实验者向他们解释说"记忆与学习"的研究涉及两个角色:老师与学习者。任务要求是:老师面对学习者读一系列的单词对,并要求学习者在看到每对单词的第一个词后,回忆其中的第二个词,学习者出错后,老师按照指示电击学习者,学习者接下来再出错,老师则加重电击强度。任务被解释完毕后,假装随机抽签,将实验者的同伴安排在学习者角色,将被试安排在老师角色。被试坐在"电击发生器"旁,控制着 30 个水平的电击,从 15 ~450 伏以 15 伏的幅度递增。在这个研究中考察被试服从率有多少。

虽然这些实验众所周知,但它们的实验设计是否存在问题,威勒和沃克教授所著的《实验设计原理:社会科学理论验证的一种路径》可以帮助你判断其实验设计的优劣。本书将依据两种实验逻辑详细阐述实验研究。第一种逻辑——理论驱动

型实验，设计是由理论决定的。这种实验的目的是检验理论，为了这一目的，研究者采用理论来设计实验。第二种逻辑——经验驱动型实验，该类实验是建立在差异法的基础上。这种实验的目的是，通过建构至少两种环境（除了一个条件有差别外，其他都尽可能彼此相同），从而发现世界中的新现象和新关系。这两种实验逻辑共同构成了促进科学知识进步的强大方法论。

译者在从事社会科学的实验研究时，发现也有不少中文书籍阐述实验研究方法，但是本书高屋建瓴，从方法论的角度来解释实验研究。从实验方法基本概念的介绍，到激发研究问题并设计实施实验研究；从经验驱动型实验分析，到理论驱动型实验分析；从如何处理研究中以人为实验对象产生的伦理道德问题，到利用计算机媒介的优势把人为因素的影响降到最小，本书涉及实验研究的方方面面，既刻画得细致入微，着眼于具体实验实例，又不拘泥于就事论事，从方法论上做了提升。相信读者通过阅读本书会对社会科学的实验研究有一个全面的了解。

译校工作从着手到交稿，历经艰辛的文字磨炼过程，当然承受这一历炼过程的不单单是我们两人，还有缪静霞、严亚利和姚鹏，他们为部分初译做了大量工作。但是，由于译者水平有限，译稿错讹之处必定存在，恳请读者批评指正。

杜伟宇　孟　琦

2009 年 10 月于上海

前　言

许多种类的研究，从现场研究(field study)到分析性调查(analytic survey)，都在追求模仿实验室实验。为什么实验室实验能成为方法论的典范呢？实验研究之所以比其他研究方法更准确，是因为它是在观察和测量的最优条件下进行的。但这并不是它成为典范的原因。实验法的完美在于它的逻辑结构。因为，与其他形式的研究不同，实验法的逻辑结构可以完全在实验室中得到实现。了解实验的逻辑是非常有用的，尤其是对于正在成长中的、努力钻研研究设计方法的实验者来说，这一点至关重要。理解任一研究如何与这种典范相背离，对于希望设计研究项目的人，或者希望批判地理解其他人设计的研究项目的人，都是很有用的。

理解实验设计的关键是，实验室实验的逻辑并不是只有一个，而是同时存在两种逻辑。本书中对这两个逻辑都加以详细阐述。对于**理论驱动型实验**(theory-driven experiment)而言，设计是由理论决定的。这种实验的目的是检验理论，为了这一目的，用理论来设计检验理论的实验。这种类型的实验已经存在了至少四个世纪之久。据我们所知，理论驱动型实验的逻辑适用于多门科学，在社会学和物理学中都是相同的。读过本书后，很少(如果有的话)有读者会同意这种说法：不同科学间存在本质差异，因此相比其他科学，实验法对于社会学而言是一种不太可靠、没多少用处的研究策略。

第二种逻辑——**经验驱动型实验**(empirically driven experiment)，建立在差异法(method of difference)的基础上。这

种实验的目的是,通过建构至少两个环境(除了一个条件有差别外,其他都尽可能彼此相同),从而发现世界中的新现象、新关系。如果由于那个差别而产生不同的结果,那么就有了发现。这种实验同样历史悠久,并且目前在不同的科学中使用这一相同的逻辑。与理论驱动型实验相比,经验驱动型实验的优点在于,人们不需要等待理论的发展来开始研究;其缺点在于,它的实验结果永远不如理论驱动型实验的结果那样值得信赖或者具有普遍性。这两种实验,分别或一起,构成了促进科学知识进步的强大方法论。

我们通过本书将实验室实验的逻辑、技术和程序介绍给更广泛的读者。一种读者是正在学习社会研究方法的人。对于他们,这是一本研究方法的教材。它通过采用大量的范例,使他们理解需要开展的实验。不过,尽管这是一本研究方法的教材,但它并不仅是教材。

对于从事研究的学者,没有什么比了解一种研究方法能做什么和不能做什么更重要的了。从事研究的学者们就是本书提到的第二种读者。关于实验法的逻辑、实验如何设计和实施以达到结果的这类知识的文献仍然是不完整的。在任何方法论文献中,都找不到设计理论驱动型实验的逻辑、对设计的解释,以及说明恰当程序的范例。如果研究者只能认识到两种实验研究中的一种类型,就会觉得实验法的广泛运用根本是个错误。例如,很多人都认为要普遍推广实验室中的实验研究是困难的,甚至是不可能的。重要的是,理论驱动型实验的结果不是普遍性的。但非常不同的程序(就是我们解释的程序)能高效地将理论驱动型实验的结果运用于实验室之外。

第 1 章概要介绍了实验法。第 2 章我们重点关注科学家如何从激发研究的问题开始,实施其研究。我们探索如何表述简单的理论,以及如何设计实验检验理论。在第 3 章,我们分析经验驱动型实验,重构这种实验的逻辑,经验驱动型实验的逻辑已经体现在米尔的准则和费希尔率先提出的统计方法中了。作为实践先导,我们将深入分析一系列实验范例。

社会学的实验与在其他科学中为其知识的爆炸式增长作出贡献的实验有什么不同吗?第 4 章展示理论驱动型实验在

许多不同科学中存在相同的逻辑，以此回答了上面的问题，并且再次通过深入分析一系列实验范例，展示研究者如何设计有效的研究。第 5 章着眼于实验者—被试的关系。我们论及了如何处理研究中以人为实验对象所产生的伦理道德问题，以及人为控制的实验环境、实验者和被试偏差。在这里，我们向读者介绍以计算机为媒介的实验及其优势，从而使相关的人为因素的影响降到最小。在第 6 章，我们展示如何将实验室环境中小的、简单的、理想化的社会系统，与实验室外大的、复杂的、杂乱的系统中的事件相关联。通过回顾理论驱动型实验，我们得出其他形式的研究如何能采用相同的逻辑形式，并得到相似的结果。

感谢斯坦福大学出版社的编辑 Kate Wahl 对本项目从初始计划到最终版本的始终不渝的支持。还要感谢国家科学基金对于本书研究分析提供的大力支持。手稿的第一稿是 Henry A. Walker 在休息日写作完成的——感谢亚里桑那大学社会与行为科学学院。感谢南卡罗莱纳大学的 Barry Markovsky, Susan H. Roschke, Brent Simpson, Shane Thye 和实验方法课程的学生，以及斯坦福出版社的评审人对本书稿提出的有价值的意见。最后，我们要对 Patrica Powell Willer 和 Joyce A. Walker 表示衷心的感谢。没有他们的耐心和支持，本书不可能完成。

戴维 · 威勒，哥伦比亚，南卡罗莱纳州
亨利 · 沃克，图森，亚里桑那州

目　录

1　什么是实验 …… 1
实验:定义 …… 2
实验法和社会科学 …… 3
三个范例实验 …… 4
　群体是如何组织起来的 …… 4
　一个有关交换的中心和利益的研究 …… 7
　现场实验 …… 10
两种实验的简史 …… 12
实验和非实验研究的对比 …… 14

2　理论和科学方法 …… 18
从问题到答案 …… 18
理论:定义和例子 …… 21
评价理论的标准 …… 24
社会学中的理论 …… 26
理论性的研究项目:从答案到新的问题 …… 27
理论研究项目的诞生 …… 28

3　经验驱动型实验 …… 32
米尔的准则和差异法 …… 33
费希尔和差异可能性 …… 36
经验驱动型实验的标准 …… 39
阿施的从众研究 …… 41

米尔格拉姆的服从实验 …… 44
社会行为的自动化 …… 48
实验室里的性别、恐惧和贪婪 …… 51
评论综述 …… 55

4 **理论驱动型实验** …… 59
理论科学和理论驱动型实验 …… 60
实验设计的原理、模型和逻辑 …… 63
验证阿基米德原理 …… 63
理论驱动型实验的准则 …… 66
将标准应用到社会学实验 …… 69
自由落体和基于地位的影响 …… 73
自由落体的研究 …… 74
基于地位影响的实验 …… 78
几何学实验:光学和交换 …… 86
应用几何光学的理论 …… 87
应用交换结构理论 …… 89
总结 …… 96

5 **实验中的社会关系** …… 98
"第一,没有伤害":对人类被试的道德及其研究 …… 99
两项研究 …… 99
人类被试的权利 …… 101
设计和开展道德研究 …… 102
实践中的道德标准:社会学研究中的欺骗 …… 103
情况说明 …… 106
招募被试 …… 109
人为因素:需求特性和实验者偏差 …… 111
需求特性 …… 112
实验者偏差 …… 114
检测和减少人为因素 …… 116
实验情境和以计算机为中介的实验 …… 118
地位特征实验的电子设备 …… 120

基元理论实验的电子设备 …… 123
总结 …… 125

6 受控制调查方式 …… 126
理论驱动型实验的效用 …… 127
规模是问题吗 …… 127
理论简单性和复杂的结构 …… 130
理论、实验和外部效度 …… 132
受控制调查的非实验方式 …… 133
实验和非实验调查的联系 …… 136

专有名称汉英对照表 …… 137

参考文献 …… 140

1 什么是实验

本书将引导你设计和实施社会科学实验。同多数人一样，我们在学习本科生和研究生的统计方法和统计推断课程的时候接触到了实验设计。不幸的是，在我们开始设计自己的实验时，几乎用不到自己所做的笔记和所读的文章与书籍。虽然同伴会提供一些建议和意见来帮助我们改进实验设计，但在多数情况下我们是通过试误(trail and error)学会如何建立科学的实验的。本书将与你分享我们通过艰难的努力取得的认识。

我们的实验设计方式和侧重点不同于其他已知的研究方法。以下三个观点是我们方法的核心：第一，科学的根本目标在于建立对现象之间关系的理论认识——任何科学都是如此。第二，理论必须在研究的锤炼中得到检验，而实验是检验理论的**最好**工具。第三，实验结果对于理论的建立至关重要。有些实验揭示出现有理论的弱点，并且指出了修正的迫切需要；有些实验则证实或支持了理论。它们的结论能鼓励研究者扩展理论的适用范围，或者使其更精确。最后，还有些实验发现了需要给予理论性解释的新现象。以上三个观点向我们表明了一个连续的过程：理论的建立激发和指导了实验设计和实施，实验的结果又促进了理论的修正和扩展，甚至有时还会产生新的理论。

实验:定义

我们在上本科生的课程时经常让学生创建假设,并且实施研究来验证假设。有些学生创建了人际关系的社会测量图,或参与受控的实验室训练;其他学生观察地位模式以及在工作或公共场所中产生的影响;还有一些学生使用政府文件来研究不同人种或种族之间的社会和人口统计特征的差异。有时他们的研究报告描述了验证"理论"的"实验"结果。这些研究通常设计和实施得都很好。然而,其中的假设不是理论,而且仅有少数的研究是实验研究。

教科书和词典给实验下了许多定义。我们使用的定义与其他从事相关研究的科学家(例如,Lederman,1993)采纳的观点相一致。

> 实验:实验就是一种探究,在探究中调查者控制感兴趣的现象,并且设置观察和测量现象的条件。[1]

实验与其他所有研究方法的区别,并不是实验与假设或理论的逻辑关系。大多数(即便不是全部)其他研究方法,都在试图仿效实验与理论和指导性假设之间的逻辑关系。然而,实验的不同在于研究者的活动,是研究者决定了研究发生的条件——进行实验研究的研究者全部或部分地创设、建立和控制研究环境。与之相反,采用**非实验**方法(nonexperimental method)(其中包括参与式观察、案例研究和大多数调查的方法)的研究者主要是在社会给定条件下进行研究。

当研究者选择了一个感兴趣的现象,并且明确了对于理解现象非常重要的条件(因素)时,设计和实施实验的过程就开始了。对于有些实验,人们需要设计假设以说明现象之间的关系,并且明确**初始条件**(initial conditions)。初始条件指的是实验开始阶段的条件,这些条件建立了实验的框架。通常,初始条件贯穿于实验全

---

1 相似的定义见于坎贝尔和斯坦利(Campbell and Stanley,1966),塞尔利茨(Selltiz,1959:94)也做过略长且相同的定义。

过程，只有在得出结论后才终止。在实验的过程中，研究者要**测量**结构、过程和路径的变化，还要记录**中间状态**(intermediate)和**终止条件**(end conditions)。研究者通过给这些条件赋予定性和定量的值来进行测量，生成数据。当研究者把初始条件介绍给参与者时，作为一个整体的实验就开始了，研究中对结构与过程的测量持续进行下去，直到测到结果时实验终止。

研究过程并没有随着测量的完成而结束。研究者还必须分析和解释数据，并将结论与指导实验设计的观点相比较。一个**好**的实验结果可以回答推进研究工作的问题，而**更好**的实验还能提出新的令人兴奋的问题，这些问题只有通过新的研究才能解决。从这个意义来说，研究是一个永无止境的旅程，在旅程的每一阶段科学家都要关心对实验检验力的提升，特别是在假设源于理论的情况下。随后我们将区分非源于理论观点的假设和源于理论观点的假设。

实验法和社会科学

怎么做才能把实验方法融入到社会学和社会科学中呢？有些人主张社会学中的实验要限定在小群体的短暂而细微的过程中。事实远非如此。官僚组织是人类建立的最大的社会结构，自从马克斯·韦伯(Weber [1918],1968)开始，这些组织由于集权而发挥着其自身的作用已众所周知。也就是说，像美国政府和天主教会这样的等级结构，之所以这样运行，是由于权力分别集中在总统和罗马教皇的手中。本书随后讨论的实验，运用权力交换网络理论来揭示像这样的官僚组织是如何通过网络结构产生集权化的。同任何一门科学一样，在社会学中，最重要的实验都是验证理论并在理论的指导下设计的。只不过理论的力量和人类想象力的局限限制了实验研究的范围。

理论和实验的结合使得研究者可以通过理论捕捉到庞杂的社会结构，并且将其分解为更小、更易控制的部分——这些部分足以成为实验调查中的客体。相反，通过理论，实验结果可以结合到越来越大的整体中，从而应用在庞大的结构中。虽

然一些社会实验关注的是小群体过程,但是与作用于更大结构的实验一样,在这样的实验中,理论也非常有价值。关于理论和实验如何结合使用毫无神秘之处。随后的几个章节将详细考察理论和实验的联系,我们将带你逐步了解理论驱动型实验的设计和实施。

本书将汲取我们所学到的实验设计和实施过程,以及理论和实验方法如何结合使用的精华,并以此作为一种强有力的方法论。如果本书引导你在下一个研究项目中考虑使用实验方法,帮助你设计成功的实验,或者帮助你评价其他人的实验设计和结果,那么这将会有助于科学的进步。如果本书能帮助你避免我们曾经的冒险与失误,那么你设计和实施实验的道路将比我们更顺利。让我们以观察一些范例实验作为旅程的起点吧。

三个范例实验

我们对实验的界定,以及关于实验应用于社会科学的评论,对于很少接触实验或科学研究的人而言,可能体会不多。什么是加深理解的最好方式?我们从爱因斯坦的话中得到启发:

> 如果你想从理论物理学家那里找出他们的研究方法,我建议你严格把握住一条原则:不要听他们说些什么,而要紧盯住他们做些什么。(Einstein,1954)

在本部分和随后的章节中,我们将采用一种经验方法:对实验设计和步骤的抽象讨论与那些验证**实际**假设和理论的**真实**实验例子互为补充。现在我们将描述三个社会实验来说明实验的各部分(初始条件、测量步骤和终止条件)如何被组织成完整的整体。这些例子表明实验的实施可以远离实验室,并且不需要复杂的计算设备。在描述了这三个范例后,我们将用它们来纠正以其他方法(如访谈和参与式观察)的标准来评判实验方法的错误成见。

群体是如何组织起来的

在这部分,我们主要关注罗伯特·贝尔斯(Bales,1950)和

其同事在哈佛所作的实验研究。贝尔斯作了开创性的研究，该研究控制了群体形成的条件，开发了群体成员互动的精确测量方法。20 世纪 40 年代后期，贝尔斯开始研究群体形成及其他基本的群体过程。起初的研究并不是验证正式的假设，相反是设计一些有关最近创立的群体的问答题目：在受控条件下它们是否形成不同的结构？如果这样，它们将会产生哪些类型的结构？为什么会形成？他们的出发点就是人们形成群体是为了完成特定目标。贝尔斯假定人们愿意自己建立（或者让别人为他们建立）互动模式，其中包括影响和身份的等级模式。这样，本没有结成群体的个体怎样聚集成为群体呢？

贝尔斯和他的同事确立了几个初始条件，研究哈佛的男性本科生（通常是大学二年级学生）。研究人员对被试进行了筛选，以确保在特定的实验中他们彼此并不熟悉。在配有单向镜的房间里，学生们坐在桌子周围。一个个标示数字的牌子被放在每个被试前面，以便于实验助手识别他们。实验者给这些年轻人一个讨论任务，然后让他们针对讨论的问题作出意见一致的决定。

其中一个讨论任务是比利・巴德的案例。这个案例是以赫尔曼・麦尔维尔的小说为基础的，比利・巴德是一名带有严重口吃的年轻海员，他是折磨、欺骗和阴谋（让他去领导叛乱）的受害者。最终，巴德打死了迫害他的长官。海军规定对殴打长官的人要处以最重为绞刑的惩罚。研究者要求实验参与者对巴德的惩罚给出一致的建议：他是应该被绞死，还是给予稍轻些的惩罚？

像比利・巴德的案例这样具有争议的问题通常可以引发人们激烈的讨论，而研究者的工作是观察讨论中显现出的互动模式。在贝尔斯和他的同事进一步形成理论，阐述群体过程中的基本问题之前，他们必须要回答这些问题：在群体互动中，研究者怎样观察？观察什么？什么被算作互动？[2]

2　许多专著和论文讨论了贝尔斯对社会理论的贡献，其中包括互动过程分析（Interaction Process Analysis）（Bales，1950）和社会互动系统（Social Interaction Systems）（Bales，1999）。按照贝尔斯（Bales，1999）的说法，对初级群体过程的早期兴趣和洞察源于对如教工会议这样自然形成的群体的观察。

贝尔斯开发了一套细致的方案,用于分类和编码互动。该系统使用一系列的数字和符号记录每一言语或非言语的行动(例如,谁作出评论,评论指向谁)。最终,贝尔斯设计了一些规则,把行为划分成 12 种,并归为三大类:①**积极的社会情绪行为**(如赞扬另一行为);②**消极的社会情绪行为**(如向另一成员表示敌意);③**任务行为**(如寻找或给出任务的方向)。

起初,贝尔斯和他的研究助手采用一个可以滚动的纸卷来实时记录行动,这种记录方法可以让人们系统地考察互动过程。一个典型的记录揭示出一个任务行为的序列:被试 1 向被试 3 寻求帮助,被试 3 随后给予被试 1 帮助。随着技术的发展,研究者开始使用录音带、录像带,最后使用计算机辅助录像记录来观察互动。

贝尔斯的研究有许多重要的发现,互动的稳定结构的发现是其中之一。在测量群体成员间参与的分配(例如,每一成员花费在交谈上的时间)时发现,多数群体只要经过 15 或 20 分钟的讨论后就会形成清晰的互动模式。在由三位男性组成的小组中,如果成员平等参与,通常可以认为每一成员讲话各占 33% 的时间。而在贝尔斯的群体中,某个人通常占去大部分讲话时间,而另一个人则说得很少。平均而言,一个由三位男性组成的小组,最积极的成员讲话占 44% 的时间,而最不积极的成员讲话占 23% 的时间。在一个由八位男性组成的小组中,如果成员间平等分配谈话时间,每人讲话将占去 12.5% 的时间。然而,数据表明最积极的成员讲话占 40% 的时间,而最不积极的成员讲话仅占 3% 的时间。该模式非常稳定。一般而言,当群体被叫回实验室做第二次讨论,同样的人将以同样的方式进行讨论。

在本书讨论的实验中,贝尔斯的早期研究是与众不同的,甚至是非常独特的,因为它们不验证假设。尽管如此,它们还是实验,因为研究者确立和积极地控制了感兴趣的现象(例如,群体的构成和大小)和实施观察的条件。此外,它们是首批建立细致测量步骤的社会实验——关于编码互动的开创性系统。它们还揭示出令人困惑的现象。一些群体似乎是在一瞬间形成结构的,好像在研究开始前就已经形成了结构。那些结构是怎样和何时形成的,为地位特征理论(Status Characteristic

Theory,SCT)的建立提供了启示。SCT 是一个解释力很强的理论,它把地位和影响过程联系起来。在我们讨论验证地位特征理论提出的假设的实验时,将会回到贝尔斯的研究。

一个有关交换的中心和利益的研究

20 世纪 40 年代,亚历克斯·贝弗列斯开始对沟通结构的变化和任务结果(像生产率、效率等)之间的关系进行研究。他研究了高度中心化的结构,像“星形”或“轮形”结构,在这些结构中某一位置可以连接到其他任一位置,其他位置却仅能连接到中心位置。低度中心化结构包括圆形——在圆形中每一位置会连接到其他两个位置上,从而形成一个封闭的链条或圆圈。在他的早期发现中,贝弗列斯(Bavelas,1950)报告说中心化结构的群体一般具有更高的生产率,完成任务速度快,并且犯更少的任务错误。

贝弗列斯的沟通结构图看上去很像在群体和正式组织中普遍存在的领导结构图:有些是高度集权的(例如,权威结构),而其他一些是更平等的(Lewin,Lippitt and White,1939)。20 世纪 50 年代中期,研究者引入图论和网络分析(Flament,1962;Harary,Norman and Cartwright,1965),使得采用量化技术来分析社会结构(表征为图表)和这些结构的位置成为可能。那些进展打开了一个全新的社会研究领域。如交换网络,在这种网络中,两个或更多的人能够交易(或者交换)商品和服务。在一个交换网络中一个人的位置是否会影响其交换的所得利益?如果会影响,那么是怎样影响的?

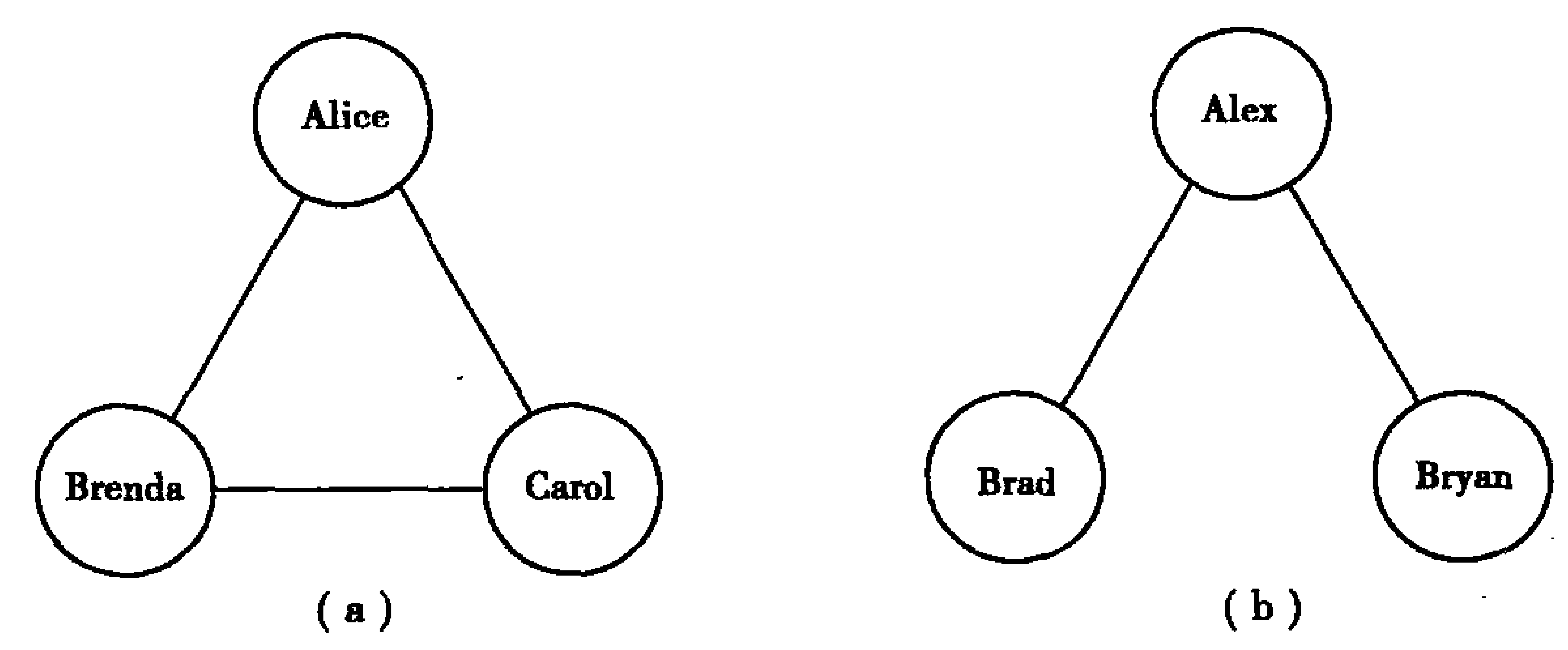


图 1.1　中心与非中心化的交换结构

图 1.1 表示了两种交换网络。第一种,三角结构(图 1.1(a))表示三个人,Alice,Brenda 和 Carol,线条表示能完成一个交换的一对成员。该网络有三个成员,每个都能与其他任何成员完成交易。图 1.1(b)有所不同。虽然 Alex,Brad 和 Bryan 还是交易网络的成员,但是在 Brad 和 Bryan 之间没有线条相连。尽管 Brad 和 Bryan 都可以与 Alex 交换,但是他们之间不能交易。图 1.1(b)是中心化的:Alex 的位置不同于其他两人,即使将所有人的名字从图中去掉,也可以从图中区分出这个位置来。相比,图1.1(a)的三个位置是一样的,一旦去掉名字,将无法区分各个位置。

像 Alex 这样的中心位置是否可能在交易网络中占有优势?它们能否比非中心位置获利更多?我们采用研究假设 1 重新叙述这个问题:

> H_W1:某一位置在每一个关系中通过交换而获得的盈利,与它在交换结构中位置的中心化程度直接相关。

中心化程度通过与某一位置连接的关系数量加以测量。因此,Alex 比他的两个伙伴处于更中心的位置,而 Alice,Brenda 和 Carol 位置的中心化程度相同。该假设认为在 Alex 和 Brad 的交换中 Alex 要比 Brad 获利更多,在 Alex 和 Bryan 的交换中 Alex 要比 Bryan 获利更多,而 Alice,Brenda 和 Carol 在交换中将获利相等。下面的实验目的在于:①了解处于交换网络的中心位置是否具有优势;②测量优势的大小。

这个实验仅关心中心化的效果,所以实验者创造了如图 1.1(a)和图 1.1(b)仅在中心化上存在差异的结构,这是合理的。实验控制能确保它们在其他方面尽可能相同。为了完成这个目标,每对交易伙伴被相同种类的标准化关系所连接。在标准化关系中的行动类似于分饼:如果一个人的饼大些,其他人的饼将会变小。两者必须在双方完全接受的分割上达成一致。假设人们在这种关系中是自利的,因为每一方都想得到尽可能多的饼,为了双方都盈利,两者都要避免意见不一致。他们的利益既对立又互补——正像自然发生的交换一样。博弈理论家称这些种类的关系为“混合动机”博弈。人们为利益而竞争(饼的份额),但又必须为赢得利益而合作。

我们的范例实验建立了一个标准关系,称为**资源库关系**

(resource-pool relation)。这不是将一张饼，而是将一组有价值的资源放在一对被试间。我们把这对被试称为 A 和 B。当且仅当 A 和 B 在分割上达成一致时，他们每个人才会得到达成一致的部分。在我们的原型关系中，库里含有 10 个资源。如果双方达成一致，认为 A 应该得到 7 个资源，那么 B 将得到 10 - 7 = 3 个资源。相反，如果双方达成一致，认为 B 应该得到 6 个资源，那么 A 应该得到 10 - 6 = 4 个资源。资源库关系就像一个交换关系，因为当 A 和 B 达成一致时，双方均获利，但是如果他们未达成一致，双方都不能获利。

实验者确立了四个初始条件。第一个初始条件是资源库关系。第二个初始条件是建立被试获取更多资源的偏好，方法是按照获得资源的数量给他们付钱。例如，每个人得到一个资源，就可获得 0.5 美元。第三个初始条件是研究者要建立如图 1.1 所示的两个结构。研究者在两种设计中使用相同的原型关系，这使得研究者易于比较在两种情境下观察到的结果。第四个初始条件明确要求在给定时间段内每一位置可以与它的每一关系交换一次。因此，Alice，Brenda 和 Carol 每个人都可以交换两次，一次与一个伙伴，跟 Alex 一样。Brad 和 Bryan 仅与 Alex 连接，每个人仅能交换一次。

实验设计必须要给出一个具体的情境设置。研究者使用了两个情境，一个是低技术情境，另一个是高技术情境。两个情境都在如图 1.1 中的三角和中心化的设计中实施[3]。在低技术情境中，被试坐在椅子上，彼此相对，为一堆筹码资源进行协商(扑克牌成为很好的筹码)。同时，用屏障把交换关系中不连接的被试分隔开来。在高技术情境中，被试坐在不同房间的计算机前操作专门软件，在他们的计算机屏幕上可以看到中心化或三角结构。屏幕显示出被试在交换结构中所处的位置与他和其他每一位置相连的位置，以及他与其他人共享的资源库关系的价值。通过控制鼠标，每一被试可以发送资源分配的提议，并且作出交换。所有被试的屏幕都会显示出每个提议和每个完成的交换。

---

3 与这里讨论的研究更接近的是 Skvoretz 和 Willer(1991)的研究。

这些结构可被作为**重复博弈**(repeated games)来研究。也就是说,每组被试都拥有几个机会,称为回合,在每一回合中被试要进行协商并且完成交换。每一回合都从放置在每对被试间的新资源开始,然后被试进行协商,达成一致,并且分割资源。当所有可能的分割均已完成,或者时间到了(通常3分钟),一个回合就结束了。该实验有几次讨论,每次讨论都招募新被试。两个实验设计都通过对成果的测量(实验的终止条件)而完成——成果指的是被试达成一致的资源库分配。在面对面场景中,由一个观察者记录那些资源分配结果。在计算机系统中,所有提议和交换自动被记录,并且存储在中心计算机的存储器中,可以随后下载。

类似于这样的实验得到的结果很清楚:相对于其他位置,在社会结构中占有中心位置没有什么优势,也不存在什么劣势。对具有同等位置的三角结构的观察发现每一交换近似5—5分割。研究者发现,在中心化结构中平均资源分配也近似5—5。此外,在高、低技术情境下得出的结果非常相似。由以上这些结果可知,研究假设1是错误的。在交换中,中心位置没有什么优势。这些结果也表明广泛得到支持的信念(称之为常识观点)即"居于中心位置是有利的",不是完全错误的,就是必须被这些结果修正的。不过,我们将随后说明,社会理论已揭示出优势中心位置的结构条件。同样重要的是,理论也说明了劣势中心位置的结构条件。

回顾前文,范例实验的初始条件包括原型资源库关系。三条资源库关系形成了三角结构,两条资源库关系形成了中心化结构。因为被试是依据交换中赢得资源的多少而获得报酬,所以实验中的稀缺环境使得实验者很合理地推论,只有从这些关系中获得的盈利才会激发被试的行动。资源库关系适合用于验证这个假设,因为资源库关系中的盈利与许多交换关系是相同的。

现场实验

现场实验类似实验室实验,因为实验者控制了研究情境中的关键因素。不过,现场实验与实验室实验也有所不同,因为

它是在现场实施，研究在此社会情境中展开，有些研究条件便不在实验者的控制之下，而是由社会情境所确定。我们描述一个由伯克、勒尼汉和罗西（Berk, Lenihan and Rossi, 1980）所实施的现场实验，该实验关注的是一个重要的实践问题。

犯罪学家证明了贫困和犯罪行为之间存在强相关关系，官方统计也表明绝大多数在狱犯在获释后很快又会重新入狱。重新入狱的比率（称为**重犯率**（recidivism rate））给刑事司法体系带来了严重问题。当许多重新犯罪的罪犯要累加上第一次定罪的刑期时，监狱的人数会迅速增加。伯克等人设计了过渡期援助研究计划（Transitional Aid Research Project, TARP）来探索制度支持（包括财政资助）对于重犯率的效果。我们把主要的假设表述为研究假设2。

H_W2：接受制度支持的狱犯比没有接受制度支持的狱犯在获释后重新入狱的可能性更小。

正如研究者在两个早期实验中所做的一样，伯克、勒尼汉和罗西确立了几个初始条件。他们从得克萨斯和乔治亚州监狱获释的男女狱犯中抽取一个样本。在劳工部的合作下，一些获释的狱犯得到了工作安置的帮助，领到了26或13周的周津贴。如果狱犯在研究期间找到了工作，他们的津贴将被削减，每得到1美元削减1美元（100%缴税），或者每得到4美元削减1美元（25%缴税）。另一组狱犯接受工作安置援助，但没有津贴。研究者也研究了其他两组获释的狱犯，一组与研究人员定期会面，但没有津贴，也没有工作安置援助；最后一组受到研究人员的监控，但研究人员不会主动地卷入到他们的生活中。总之，研究跟踪以下6组狱犯：

1. 接受工作安置援助和26周的津贴，按所得工资100%缴税[4]。

2. 接受工作安置援助和13周的津贴，按所得工资100%缴税。

---

4 津贴被限定在各州的失业救济上。从乔治亚州监狱获释的狱犯得到70美元。在薪金被减少前，乔治亚州监狱获释的这些人能从工资中得到额外的13～75美元。得克萨斯州监狱获释的这些人每周得到63美元，在救济被减少前，能得到额外的8美元。

3. 接受工作安置援助和 13 周的津贴,按所得工资 25% 缴税。

4. 接受工作安置援助。

5. 接受研究人员的定期会面。

6. 接受研究人员非直接参与的监视。

TARP 的结论在学术研究者和政策制订者中引发了大量讨论,但是结论是明确的:给获释狱犯津贴没有明显地减少得克萨斯或乔治亚州出狱犯的重犯率。在获释后一年内,几乎一半(50%)的男性罪犯由于新的罪行重新入狱。研究中获释女性的数量太少了,很难让研究者针对女性罪犯完成细致的统计分析。尽管如此,得克萨斯和乔治亚州监狱获释的男性罪犯的数据几乎是一致的。数据不能支持研究假设 2。

两种实验的简史

三个范例实验表明了社会科学实验的范围。贝尔斯的研究关注微观社会过程,集权实验对小规模的和大规模的结构有意义(例如,官僚组织),TARP 的研究结论对整个社会具有意义。除了这些大量观察,本书的核心是作出一个区分:在科学中有两种实验——**经验驱动型实验**和**理论驱动型实验**。区分这两种类型的实验是非常重要的:经验驱动型实验**探索或发现现象**,而理论驱动型实验**验证理论**(Hempel,1966)。本章这三个范例实验都是经验驱动型实验。第 2 章将呈现理论驱动型实验设计。

理论驱动型实验如果追溯到阿基米德,有 22 个世纪以上的历史了。公元前 230 年,阿基米德著书提出杠杆原理,并且似乎系统地验证了该原理(Archimedes[公元前 230],1897)。不管阿基米德是否作过验证,可以确定的是,伴以数学分析的理论驱动型实验的**现代**传统可追溯到伽利略对自由落体理论的验证。虽然伽利略的著作早在 1607 年就已完成,但是天主教会将其推迟到 1636 年出版。因此,目前理论驱动型实验的

现代传统已有4个世纪了(Galilei [1636],1954)。第4章将分析阿基米德和伽利略的实验。

从伽利略开始,理论驱动型实验和理论发展相互影响,持续贯穿整个科学,直至今天。1664年,牛顿接触到笛卡尔的解析几何后,出版了《数学原理》(Newton[1686],1966),完善和拓展了伽利略的力学。从那以后,首先在物理学,然后在化学,近期在生物学,实验和理论发展相连接的研究潮流再次兴发壮大。虽然在社会学中主流方法论几乎或根本不能归功于这些发展,但是正如我们在第4章所阐述的,社会学中的理论发展与研究却是与物理科学的工作主流相一致的。

到了19世纪中叶,现代科学实验方法已经确立起来。大家非常明确,有种种原因可以认定实验方法能扩展到新近出现的社会科学中。然而,这种扩展没有实现。米尔的颇具影响力的著作《逻辑体系》(Mill [1843],1967)解释了科学是如何运作的。然而,理论和研究之间的关系,以及两者如何相互作用来增加知识,在其著作中并没有论述。相反,米尔将实验奉为发现经验规律的最佳工具。今天我们把经验主义视为这样一种观点,即理论和定律是描述规律的表述。经验主义的产生更多归功于米尔,虽然他陷入了相信科学的目的是寻求不变规律的错误之中[5]。米尔以及其他追随其后的经验主义者和统计学家中最著名的费希尔,将经验主义方法规则化,后来这种方法被大多数社会学家所采用。

现在具有讽刺意味的是,从米尔至今160多年,经验主义者的方法虽然扎根于错误理解科学的基础上,但是已逐渐完善成为强有力的研究工具——这些研究工具不能产生理论,也不能将社会学提升为一门理论科学。尽管如此,由米尔提出,费希尔随后进行完善的这种方法,仍被广泛用于发现相关规律,不仅用在社会学,还用在医学、农学、心理学领域的

5 正如理论驱动型实验是在伽利略之前进行的那样,经验主义者的实验也早于米尔。培根(Bacon[1620]2000)(伽利略同时代的)就表达了米尔的观点。然而,米尔在观点的系统表述上更复杂,并且对现今的科学实践更具影响力。

社会心理学,以及其他许多领域。对于任何一个很少提出理论的领域,没有比米尔和费希尔的实验方法能更好地洞悉世界的了。在第3章我们将讨论经验驱动型实验设计,揭示出它们是如何遵循米尔和费希尔的逻辑建立起来的。在这一点上我们与米尔能够达成一致:他的这类实验是很好的发现方法。另外一方面,理论驱动型实验是验证理论和促进科学理解的最好方法。

实验和非实验研究的对比

实验是验证理论的最好方法,因为它们比起其他研究方法拥有许多优势:①它们让研究者控制一些或所有条件,在这些条件下一些现象能被观察到;②它们对感兴趣的因素和过程作更精确的测量;③它们允许研究者创设和再现必要的初始条件来验证由理论得出的假设;④有效地改善理论、假设和研究发现之间的匹配程度。所有这些优势使得实验方法拥有强大的检验力。因此,产生积极结论的实验强有力地支持了假设,产生消极结论的实验也可以同样有力地拒绝假设。出于这些原因,实验比其他数据收集技术能更好地验证假设。既然实验研究有这样和那样的优势,客观的观察者就要问了,为什么在社会学中很少有研究使用实验方法?

我们从三方面回答这个问题:第一,社会学家中很少有人接受过专门关于实验的,或者如何设计和实施实验的教育;第二,社会学家很少有学习实验技术的动机,因为社会学家普遍认为实验不能产生有用的社会知识;第三,有些实验会引起尴尬的伦理问题,这样在研究开始前必须花大量时间完成征得人类观察对象同意的过程。

本书旨在通过详细阐释经验驱动型和理论驱动型实验的逻辑,以及两种类型中的优秀设计的原理来解决第一个问题。本书也讨论一些尴尬的伦理问题及如何解决它们。后面的章节会论述这些问题。这里我们面临这样的信念挑战,即社会学

科不适合实验研究。

怀疑社会实验效用的研究者认为原因有这样几个:第一,许多人指出实验的控制条件是人为设置的,因而,没有理由认为人们在实验情境外也会遇到相似条件。第二,实验被试知道自己在做实验,被观察。相比较而言,被研究的人们在开始日常活动时(在有些案例研究和某些形式的参与式观察中)不需要意识到他们是被观察的对象。感知到自己受到观察似乎会对行为产生实在而未知的影响。第三,由于抽样方法是不明确的或者不清楚的,实验被试通常来自于定义不清楚的总体[6],于是从实验被试推广到更大的总体(这在调查研究者中是一个普遍的惯例)是不可能的。

我们能容易地说明这种推论是被另一类似观点(即调查研究不能产生有用的知识)所误导的——即使社会学者和其他社会科学家毫无疑问是调查方法最老练的使用者。这一观点认为,多数调查是在社会现有条件下实施的,而这些条件不能为调查者所控制。缺乏实验控制,结论就不能有力地支持假设,或者不能得出明确拒绝的结论。虽然不是不可能,但精确测量确实很难在现场情境中实施。普遍化结论受限于从样本到整体的归纳,而每一整体受到过去的特定条件、时间的限制。由于调查者仅能对过去提出普遍化结论,因而无法解释当前事件和预测未来事件。由于研究者不能获得一般化的、普遍的知识,所以从调查中所获得的知识解释或预测力很弱,或几乎没有。

以上两种评论都被误导了。对实验的评论实际上表明实验方法是有缺点的,因为它设计了糟糕的调查。对调查的评论也是类似的:调查方法也有缺点,因为它设计了糟糕的实验。事实上,两种方法逻辑明显不同,有不同的目的,不应采用评价其他方法的标准来评判它们。其他方法也是如此,像案例研究

6 例如,考虑以在某大学注册心理学导论课程的所有学生为总体,要求学生参与5个可能实验的任意3个。根据我们的经验,参与者中大多数将是新生和大二学生,他们当中物理和工程专业学生所占的比例很小。

和历史比较研究。每一方法都有自己的优势和弱点,都不应该按照实验或调查的研究标准来评判。如果一种方法被判断为最适合的方法,那么它通常对某些问题是最适合的方法,而对其他问题则不如其他方法有效,甚至整体上是无效的,因此根本不能使用这种方法。例如,实验方法最适合于理论验证,但是不能用于发现群体中流行的观点。

其实,根本不存在最优的研究方法。例如,要了解今天的街角社会,调查、实验或者历史调研都不能搜集到想要的信息,所需要的就是案例研究,还可能是一系列的案例研究,且毫无疑问是由参与式观察者来实施的。回顾有关交换和中心化的实验,可以想象现场研究能够考察盈利和交换中心之间的关系,但效果并不好。我们不清楚,研究者怎样才能对中心化交换结构和全连接三角结构进行取样,让两者足够相似以进行准确比较。此外,前文介绍的实验结构建立在实验参与者完全被隔离的基础上。相反,作现场研究的研究者没有设计任何检验方案,以确定在现场中所研究的结构是否受到其他邻近结构的影响。

因为社会研究方法是出于不同目的被开发出来的,所以很少相互对抗,而常常互补。调查要让某一公民样本回答问题:"在交换结构中处于中心位置是否能说明其具有优势?"在分析回答时,就算某些被调查者没有能力回答问题也并不重要:调查关注的不是知识,而是观点。调查能用于发现"中心位置具有优势"的**信念**是否广为流传。信念是否广为流传对于社会学家、经济学家和消费者保护团体具有十分重要的意义。考虑这样一个常被证明的结论:特定种族的少数群体成员会比多数成员为汽车支付更多的钱(交换结果),可能是因为很多人持有汽车经销商具有优势的不正确信念。如果调查发现不正确的信念广为流传,那么实验能研究这些信念对交换的影响。

实验方法不可能首先提出有关任一群体成员所持有观点的问题,但是它能成功地研究重要的社会现象。某些人是如何对他人行使权利的?权利是如何被创造的和应该如何被测量?

权利是否能被抵消掉？在什么条件下人们遵从其他人所表达的观点，在什么条件下人们不遵从？影响力是否随地位而发生变化，以至于高地位的人比低地位的人更具影响力？影响力是否能转化为权利？为什么有些人愿意服从权威，而有些人拒绝服从？这些都是适用于实验法的研究问题，而且对于这类问题实验法毫无疑问是最适合的方法。此外，这些研究问题在最近已经采用实验方法进行了研究，并且社会科学家已经运用这些实验结果解释和预测实验室外的现象了。

实验方法是用于发现现象和验证理论的重要工具。实验和理论之间的关系是本书的核心。这里已经介绍了实验方法，接下来我们将转为阐述理论和与实验间的关系。

2 理论和科学方法

科学的首要目的是对现象进行描述、分类和解释。科学以问题为开端,要取得进步,就必须发展和验证理论来回答那些问题。本章以提出划分科学问题的三类方案为开始。对这三类问题的解答虽然有着不同的科学目的,但研究者把这三类方案都作为推动理论发展的整体策略的一部分。接下来,我们将给出理论的定义,并用一个例子说明如何正式地用一般语言表达命题。我们将描述理论的若干形式,并且确定评价理论的标准。最后,我们将说明怎样结合问题、理论和理论的实验验证来创建研究项目,从而推动理论的发展。

从问题到答案

每个科学研究都是从问题开始的,然而,正如第1章我们介绍的三个实验显示的,问题可能是千差万别的。问题的实质性差异,远不如提问所采取的形式和相应产生的各种答案重要。我们重新回味那些实验,并不是为了重申那些显而易见的差异,而是为了探讨引发出那些实验的问题之间的重要差别。

沃克(Walker,2002)确立了普遍的三类或三种科学问题。第一类问题探寻某些现象的存在及其特征,第二类问题探寻现象之间的关系,最后一类问题寻求两种(或更多)现象间联系的原因。

科学家探寻的第一类问题就是形式更复杂的“孩子式提问”:这是什么,y? 某一特定的 y 与另一 y 如何相似或不同? 类似的这些问题是所有科学探究的出发点。第一类问题

(Walker,2002)激发研究者对现象进行系统而细致的观察。这些深入的观察促进了对现象的严谨描述和系统分类的发展。第1章贝尔斯(Bales,1950)的研究项目中的早期实验研究了第一类问题。贝尔斯通过确立各种言语和非言语行为,以及开发一个普遍的分类系统,回答了很多第一类问题。

第二种问题(第二类问题)探寻:哪些因素与事件 y 的发生有关,或者与事件类 Y 的部分(例如,所有与 y 相似事件的集合)相关?[1]因为他们首先关注的是关系,然后才是现象,因此第二类问题比第一类问题要求更复杂的描述。一旦贝尔斯和他的助手建立起一套对行为进行分类的可靠系统,就能确立不同类型行为间的模式和共变关系。例如,贝尔斯确立了小组成员的评价和谈话所花费时间之间的关系,以及其他成员如何评价他们观点的质量。对于第二类问题的回答有时也能激发研究者提出额外的问题,以寻求解释。某些寻求解释的问题是第二类问题的变式,它们引导研究者发展经验主义或统计的解释。但仅有一类问题激发研究者去发展理论上的解释——我们称之为第三类问题。我们将在讨论经验主义的解释后,再继续讨论第三类问题。

经验主义的解释回答了一种或多种的具体事件如何关联到(或解释)另一个具体事件或其中变量。它们关注于具体的、特定时间的事件或因素(xs 或 Xs),这些事件或因素导致(或引发)了研究(y or Y)[2]中的现象。社会科学家擅长设计**定性**和**定量**的经验主义解释。

---

1 我们使用小写字母符号(如 x 或 y)来表征具体事件或现象。苏茜嫁给约翰是一个具体的事件。我们使用大写字母符号(如 X 和 Y)代表事件类。事件类是同一类型事件的集合,例如南加州和亚利桑那州2006年核准的所有婚礼。按照威勒和韦伯斯特(Willer and Webster,1970)的说法,我们使用这些术语表示一些观察到的现象(如变量或测量概念)。有时(见后文),我们将使用 X_C 和 Y_C 表示理论建构,以及不受时间和地点限制的一般术语。在第4章我们将讨论理论构建,比如分别从物理和社会科学中得出的"自由落体"和"地位"。

2 对于我们所称的经验解释,沃克(Walker,2003)使用术语"历史解释(historical explanation)",因为开发和使用如此解释的那些人采用了历史的方法产生知识(Berger, Zelditch and Anderson,1972:ix)。科学不能超越观察和观察中的推论而延续下去,这种观点就被称之为"经验主义"。我们不是经验主义者,因为我们认为观察和推论不是尽头,而是朝着理论建立和验证迈进了一步。

定性的经验主义解释关注于解释一个独特的事件——y。例如,韦伯(Weber [1904],1958)试图解释西方禁欲主义的新教教义对于资本主义崛起的影响,它本身就是一个独特的事件。在2001年的"911事件"以后,政策分析家们聚焦在可能对美国发起的种种袭击上,他们首要的目标是确定导致每次袭击的具体的连锁事件(xs)。分析者发现①很多攻击者都是沙特阿拉伯民族;②他们合法地进入美国,然后"消失"在美国社会中;③他们只付单程票;等等。这样分析的目的是为了寻找是否存在政府或执法官对特定的x(或xs)进行干预以阻止袭击发生的办法。

定量的经验主义解释关注于一个事件类——Y。更加准确地说,研究者通过确定与其他事件类(Xs)中现象的共变量,来解释组成这个事件类的现象中的变量。例如,利希特、麦克劳克林和瑞巴(Lichter, McLaughlin and Ribar, 1997)对"为什么有些女性是家族里唯一的首领,而有些不是"感兴趣。全国范围的数据分析表明13个自变量(Xs)与女性首领家族的发生率(因变量)存在显著相关。这些变量包括女性的收入、人口数、男女性别比例,等等。他们的初始模型能够解释Y(女性首领家族的发生率)中83%的变异($R^2=0.834$)[3]。分析家们运用多变量技术来解释因变量的变异,着手于定量的经验主义解释。我们将在第3章中讨论基于统计设计的实验。

回到第1章讨论的实验,中心位置的研究及关于经济支持和累犯的TARP*研究反映了第二类问题。每个研究都包含着对现象间因果联系的推断:研究交流网络的人们已经观察到位置的中心性关系着对资源的控制,在此基础上他们推测中心性也关系到交换网络中获利的分配。相似的,犯罪学家和社会政策研究者也认识到贫穷和犯罪行为是相关的。研究者设计TARP的研究,就是用来发现增加被释放囚犯的财政来源是否可以减少重犯率。

---

3 R^2(读作"R的平方")是指由自变量变异所引发或解释的因变量Y的变异量。

* 伯克,勒尼汉和罗西(Berk, Lenihan and Rossi, 1980)设计了过渡期援助研究计划(TARP)来探索制度支持对于重犯率的效果——译者注。

促进理论发展的第三类问题与寻求经验解释的第二类问题形成鲜明的对比，关注诸如 X_C 和 Y_C（例如，社会地位和影响力）的理论概念之间的关系，只能通过设计包含说明这种关系的一般机制的理论加以解答。促进理论发展的许多问题的解答历程也包含了第一类问题（y 是什么）和第二类问题（哪个 Xs 与 Y 有关系）的答案。第 1 章中描述的实验没有一个是能反映理论概念之间关系的第三类问题，也没有一个是被设计用来验证理论的。

因为理论发展是科学进步的核心，实验是验证理论的最好方法，所以理论和实验之间的关系是本书的核心。理论驱动型实验的新手可能会问："我从哪里找到设计实验的理论呢？"我们推荐两个来源。首先，研究者可以依据现有的理论工作——如果研究者走这条路，其首要任务就是探索没有被研究的推论，或者从理论中作出新的推论。其次，研究者可以创造新的理论，走这条路的人会发现他们将经历上述三类问题。在任何一种情况下，成功取决于对理论的理解和利用。因此，我们现在转向回答这个问题：什么是理论？

理论：定义和例子

理论解释现象之间的关系[4]。我们所遵循的这一定义来自最受推崇的理论构建文本（Cohen，1989：178）的观点。

> 理论：理论是一组相互联系的、普适的命题，命题中所蕴含的一套规则和程序可用来产生新的命题。

我们通过强调理论必须包含相互关联的命题来揭示出这个定义。相互关联的命题是推演新命题、解释研究中关系的必要条件。现在我们描述一个理论来回答这个问题：在交换网络中为什么位置中心化和获利成正相关？注意问题表述的形式：为什么一种现象（位置中心化）以特定的方式（正向的）与第二种现象（获利）相关？将下列三条论点或理论命题（TSS）称为

4 更准确地说，理论是解释关于现象间关系的命题（Cohen，1989）。

位置中心化的理论:

TS_1:在交换网络中位置**中心化**与对其他位置的**依赖性**负相关,越非中心化的位置越有依赖性。

TS_2:在交换网络中位置的**依赖性**与其**权力**负相关。

TS_3:在交换网络中位置的**权力**与**获利**正相关。

这三个命题是相互关联的,因为每条命题与另外至少一条命题共有同一术语(见黑体字术语)。

理论表述也必须具有普适性,不能指明具体的时间、地点或者事物。在上面的理论表述中黑体字表示的关键术语和理论建构没有一个指明具体的时间或地点。事实上,一般的理论阐述经常取代自然语言的理论阐述,例如上面给出的位置中心化理论,可以用纯粹的符号语言表达成下面的形式(例如下面的表达式中就有正体、斜体的分别)。

TS_{1r}. $A_C R_1 B_C$

TS_{2r}. $B_C R_1 C_C$

TS_{3r}. $C_C R_2 D_C$

"A_C 位于特定的与 B_C 的关系 R_1 中",我们这样来解释命题1。对于命题2,我们写道"B_C 位于特定的与 C_C 的关系 R_1 中",这同样适用于命题3。命题1,2,3 中的术语在代数意义上具有普适性。这些我们用来表示关系的符号 A_C,B_C,C_C,D_C 和 R,都能够用于表征任何现象或关系的理论概念。我们通过对 Rs 加注脚来表示在同一个理论中,一条命题描述的关系 R 不同于另一条命题中的关系 R。作为一个演示,让你自己相信诸如"中心化"、"依赖性"、"正相关"之类的理论概念可以被命题中如 $TS_{1r\sim3r}$的符号替代。

理论命题要相互关联并且受普适性术语限制,这一要求对理论验证具有重要意义,相互关联的论点使得科学家们创造出全新的命题——这是理论与孤立的假设相区别的本质。例如,逻辑的规则可以应用到上面的理论中,来推导出我们标示为推论1(D_1)的命题。注意推论 1 也可以以纯符号术语表示为 $A_C R_x D_C$。

D_1:位置的**中心性**与其**权力**正相关(参考 TS_1 和 TS_2)。

接下来我们可以立即给出另外两个推论,标示为推论2 和

推论 3。同上，这些推论也在逻辑上遵循命题 1～3。

D_2：交换网络中位置的**中心化**与**获利**正相关（参考 TS_1，TS_2 和 TS_3）。

D_3：位置的**依赖性**与**获利**负相关（参考 TS_2 和 TS_3）。

有了位置中心化理论，就要考虑如何验证它。在验证过程中，我们将创造一些条件，在这些条件下的观察能评估理论中一个或多个命题。研究者们使用观察结果来评价理论中普适性的价值，这似乎看起来是个悖论，因为观察结果**总是**指明特定的时间、地点、事件。科学家们如何将不指明特定时间、地点或事件的理论与具体的观察联系起来呢？这一悖论已被解决，因为理论有产生新命题的能力。

在将普适的、抽象的事物与特定的、可观察的事物联系起来的新命题产生之后，研究者再来验证理论。这些命题被称为**连结命题**（linking statements，LS）。因为它们连结了两种术语，一种是抽象的、理论的，另一种是具体的、经验的（Cohen，1989：82）。如下所述，经验主义的术语指出了测量方法，我们用位置中心化理论和第 1 章中所描述的实验，来表明连结是如何完成的。

LS_1：以在交换网络中与某个位置相连的关系数目来测量其中心化程度（初始条件）。

LS_2：以在资源库分配中得到的资源（扑克牌、筹码等）数目来测量交换中的获利（终止条件）。

像 LS_1 和 LS_2 这样的连结命题允许研究者用具体的变量替换理论概念，并用具体可观察的术语推论出新命题。包含可观察的术语（变量）的推论通常被称为假设，从而与那些仅包含理论概念的推论命题相区别。假设 1（H_1）就是新命题的一个例子，该命题是利用初始条件命题和结果条件命题创设出来的，从而用来解释位置中心化理论。注意 H_1 用具体的语言重述了上面的 D_2。H_1 与第 1 章中第一个研究假设 H_W1 具有同样的关系。

H_1：与一个位置连接的关系数目与其在资源库分配中得到的扑克、筹码的数目**正**相关。

回到第 1 章中我们描述的实验，Alice，Brenda 和 Carol 都同

样各有 2 个连结,在每个关系中他们都得到同样数目的资源(大约 5 个),至于中心化的结构,Alex 有 2 个连结,比各有一个连结的 Brad 和 Byran 要多,然而在这个中心化网络中,这三个人在每个关系中仍然获得相同数目的资源(大约 5 个)。

我们对于理论及其经验主义解释的讨论开始于一个理论例子(位置中心化理论),该理论被我们表述为一套三个相关的普适性命题。我们已揭示出这三个命题的关系如何使得我们推导出新命题。我们也揭示了普适性命题如何被赋予经验主义的解释,从而使其与实验得出的观察结果相联系。回顾第 1 章中讨论的实验,我们现在可以得出中心化结构的数据与 H_1 不一致,也就是说,H_1 是错误的[5]。因为它是错误的,逻辑也要求研究者将其普适相当的推论 2 视为错误的。事实上,数据证伪了中心化理论。

然而,我们不能因理论的证伪而失望,因为长期的经验教会我们不要把否定的结果与知识发展的失败相混淆。发展理论的研究项目可能会陷入死胡同,很多被证伪。通常这些死胡同是学习的经验。现在我们知道位于中心并不能自主产生权力的差异。我们也知道,在实验室之外,位于中心位置的人们通常很有权力。因为我们发展了对网络中权力的知识,我们现在知道,为了解释这些中心化位置的权力,我们必须超越中心化,发现产生权力的条件。我们将在后续的章节中揭示出一些发现。

评价理论的标准

若干标准对于理论的评价很有用。它们包括①逻辑的一致性;②范围的广度;③经验的支持。一个理论若包括或暗含命题"中心化和依赖性负相关"和"中心化和依赖性正相关",则这一理论就不具有逻辑一致性。这两个命题相互矛盾。有

5 如果一个或几个单独的观察发现了反面的证据,那么一个假设或者理论可以被宣布为"暂时性错误"。证伪需要可重复性(见本章随后所讨论的重复性对证伪的重要意义)。

时研究者可以通过细致地考察理论的内部结构而发现其中的不一致。更通常的情况下，当理论被应用及其预测不合逻辑时，研究者会发现其中的不一致。在任何一种情况下，如果可以的话，当人们发现理论包含或产生不一致时，理论的创造者就有责任消除它。直到他们这样做了，理论才会变得有效。

理论通常应用于指定的范围条件。**范围条件**总是以普适性的非具体术语方式表述。例如，理论经济学——经济学家称之为新古典主义的微观经济学，其范围限定于市场经济。市场经济这个术语是一个普适性的术语，因为它不牵涉特定的时间、地点或经济布局。古代中国和古罗马（Weber [1918]，1968:164）可能包含市场经济，大多数当代社会包含市场经济，未来将出现的很多社会无疑也包含市场经济。“当今的所有经济”和“20 世纪的美国经济”的表述不能界定理论经济学的范围，因为这些术语不具有普适性。它们涉及特定的经济和特定的时间、地点。

更普遍的是，范围条件界定了两个领域，一个是理论应用的领域，另一个——通常更广——是理论未应用的领域。虽然理论是普适性的，但还存在不能应用的一般条件。当理论阐释恰当时，范围条件将会排除这些应用。因此，市场经济作为一个范围条件，会排除新古典主义的微观经济学在公社社会的应用，像乐队和部落。这个排除是很重要的，供应和需求在没有市场的经济中也变化着，但是没有一个会影响到价格，因为价格根本不存在。任何把新古典主义的微观经济学应用到这类社会的尝试都是没有意义的。此外，找不到供求关系对价格的影响并不意味着证明该理论是错误的，因为无市场的社会是在这个范围之外的。

第三个要考虑的是理论博得经验支持的程度。一个结构良好却未得到经验支持的理论可以产生推论。然而，理论没有经验支持，就不能使其推论成为解释或可靠的预测。难道一个理论的预测和解释就限定在被验证的现象的范围内吗？根本不是这样！理论从来不被之前的验证所限制，相反，理论的目标是要超越已知的范畴。然而，理论更广泛的解释和预测的影响力与其现有验证的成功有关。

理论的经验支持一定达不到证明为真的目标。例如,虽然有研究支持假设 1(或者第 1 章中的研究假设 1),但仅凭那个研究并不能证明假设是真的。为了做到这些,我们不得不研究存在于过去、现在的,及将来一直会存在的所有中心化结构。然而,就像我们所说明的那样,这个假设被证明是错误的。任何理论都可能会出现同样的情况。事实上,上面描述的一些例子表明研究假设 1、假设 1 和位置中心化理论是错误的。

尽管证据表明位置中心化理论是错误的,但还必须要加上两个重要的限制。首先,证伪不能仅依赖于几个孤立的观测结果,还要不断地重复。重复验证的能力恰恰是理论驱动型实验的强项,而不能重复也正是社会学中其他调查方法的弱项。例如,早先讨论的交换和中心化实验很容易被阅读这本书的任何人复制或扩展到其他结构中——至少在低技术情况下可以做到。事实上,任何读者使用因特网上可获得的软件,也可以重复这个实验的高技术版本。在本书的后续章节,我们将会说明如何获取那些软件。

关于证伪还有一点必须说明。有时,理论只有在更好的新理论出现并被取代的时候,才会被认为是错误的。是否需要更好的理论,或者是否只需要反证的证据,其实是这样一个问题:"如果理论被抛弃,损失的会是什么?"位置中心化理论能单单被证据所揭破,是因为它既没有先前的经验支持,也没有超越其检验的经验推论。相比而言,如果任何理论具有大量的先前经验支持和超越其被检验的大量推论,那么该理论就是一个值得考虑的有价值的理论。与其在发现反证证据后放弃所有的理论预测力和解释力,不如归纳出新的范围条件,包括理论有效的现象和排除理论无效的现象。然后只有等到一个适用范围更广的新理论出现,当前的理论才会被认为是错误的。

社会学中的理论

如果社会学中的理论都被形式化,以至于假设的验证轻而易举,那么实验者的任务就会变得直截了当。一些理论以规范

的语言形式书写，采用逻辑或数学语言，从而达到精确的目的。对于那些理论，实验者提取假设的任务就变得更直接了。正如我们在后续章节中所表明的那样，这样的理论在开发验证它们的实验设计中也很有用。

在社会科学中，理论采取多种形式。就像马克思和韦伯的经典理论一样，它们最大的用武之地是建立在对历史和当今社会的调查基础上的。诸如此类的理论将成为有价值的资源，不仅仅在于其假设，更在于那些形式主义和实验构建的社会概念。后续章节将说明对社会的经典主义理解的正式理论。一些社会科学理论混淆了世界和政治意识形态的命题——马克思就是这样的。在这些情况下，研究者会面临挑战——要从不能检验的意识形态中分离出可以检验的理论部分。还有些人的工作集中在那些被称为“元理论（metatheory）”的东西上面。元理论包含构成理论基础的预想。因为预想是不可验证的，所以在这里我们不关注元理论。

理论性的研究项目：从答案到新的问题

很多社会科学家建立或致力于研究项目。也就是说，他们计划并执行一系列相关研究，每个研究都增加了先前获得的知识。我们需要区分两类研究项目。**积累式的研究项目**（cumulative research program）是一系列相关的研究，系列中的每个研究都能提高研究者“辨识和解决问题的能力”（Cohen，1989：293）。像贝尔斯（Bales，1950）所做的研究项目一样，累积式的研究项目丰富了描述、分类和经验解释的知识仓库。然而，开展最好的、最先进的科学产生和维持**理论性的研究项目**（theoretical research programs）（Wagner and Berger，1985）。理论性的研究项目是由一个理论或一系列相关理论，以及验证这些理论的研究组成的。这些项目也可能包括将理论知识应用到具体情境和问题的研究（Wagner and Berger，1985：705）。

成功的累积式研究项目能引发理论研究项目。研究者可以依次采用类型1，2和3的问题以及他们发现的答案，从而建

立先是累积性的,然后是理论性的研究项目。我们采用贝尔斯的研究来解释这个过程。正如第 1 章所描述的,贝尔斯用他早期的观察结果开发了一个用于描述和分类互动的系统。他的研究提出并解答了一系列第一类型的问题。什么是"行动"?讨论进行到1 分30 秒时被试1 直接对被试2 做出了哪种行动?这种行动与被试4 直接对整个小组在4 分11 秒时的行动有什么异同?然而,贝尔斯关于记录和编码互动的系统是一个重要的开发,其最重大的意义在于,它使人们有能力系统地观察行为,并提出更加复杂的第二类问题。

正如已经提到的,贝尔斯描述互动的系统使他观察到,花费在交谈上的时间不是平均地分配给任务小组的成员。我们可以设想研究团队问下列问题:"什么东西(或因素)与组员们花费在谈话上的时间相关?"这是第二类问题。现在,我们知道了组员们会感知到,谁对最好的意见作出贡献,谁尽最大努力维持小组针对任务的工作,谁是他们最喜欢的,这些方面的感知都与花费在谈话上的时间正相关。然而,如果早期研究并没有回答我们所描述的第一类问题,那么他们就不能发现这些答案。太多的项目中止在这一水平上(例如,累积式研究项目),但是贝尔斯紧接着就创建了理论研究项目,许多他以前的研究助手创建了其他的但与之相关的理论研究项目。约瑟夫·伯杰和他的同事就创建和扶植了这样一个项目。

理论研究项目的诞生

1950 年代早期约瑟夫·伯杰与几个年轻的研究生助手在贝尔斯的实验室工作,一起继续发展期望状态理论,地位特征和期望状态相关的理论。我们现在就来描述贝尔斯研究中的行为模式和促使伯杰发展地位特征理论的第三类问题(Berger, Fisek, Norman and Zelditch, 1977)。

一个典型的贝尔斯小组在大约 20 分钟内产生一个清晰可识别的互动结构。然而,贝尔斯的助手注意到了一个重要的异常现象:一些小组几乎瞬时产生了清晰界定的互动结构,好像

他们在进入实验室前就已经有一个小组结构。这是如何发生的呢?

贝尔斯很小心地把完全陌生的人分配到实验小组。我们在这项研究中列出的一个初始条件,就是这些人与其他被试相互间不熟悉。在系统性地复查了他们的研究记录和数据后,贝尔斯的研究小组发现这些在实验室的等候室里匆匆拼凑的完全陌生的人,做了通常陌生人聚集在一个陌生地点要做的事:他们彼此交谈。

等候中的陌生人和坐在他们附近的人谈话,很快就知道"汤姆·琼斯"是大二学生,没有决定好专业,并且居住在罗维尔舍院——正是最后的这条信息至关重要。哈佛学院里的高年级学生分住在12个舍院——在其他学院或大学通常被称为宿舍或居住楼。房子按大学生所共知的社会地位分等,有些舍院地位高,有些地位低,被试之间的社会地位受到他们居住舍院等级的影响。住在很有声望的舍院的学生比其他在相对差点的舍院里的学生谈话要多,贝尔斯和他的研究助手揭示了第二类问题的答案:什么因素与组员在谈话上花费的时间有关?其中一个答案就是居住舍院的等级。

这个故事很有寓意,实验者必须警觉,这是因为他们不能控制的条件却能无意地(并且认识不到)产生了他们正研究的现象。贝尔斯避免招募来自同一所房子的学生,因为他想使自己的每个实验在学生彼此陌生的情况下开展。实验控制保证每个实验小组的被试都来自不同的房子。然而,一旦被试交换了他们的住宿信息,舍院阶层的差异就影响了小组的互动,并产生了可观察到的快速发展的结构。其他小组同样由来自分级居住系统的被试组成,也会发展出同样的结构吗?他们的小组结构会仅仅因为在实验中被试发现他们的居住状态的差异(或者其他地位的差异)比较晚而产生得比较缓慢吗?我们不知道贝尔斯的研究是否能回答这些问题,但我们知道现今的一个重复研究可以回答它们。

约瑟夫·伯杰认识到,这一现象的普遍性要超出哈佛的舍院和大学生。来自高地位群体的人在由许多不同地位背景的新人组成的小组里通常拥有较高的地位。一个经典的例子就

是陪审团的主席一般比其他陪审团成员拥有更高地位的工作(Strodtbeck,James and Hawkins,1957)。伯杰提出了第三类问题(见 Berger, Cohen and Zelditch, 1966, 1972; Berger, Fisek, Norman and Zelditch,1977):

> 什么能解释人们在小组外的社会地位与小组内的地位、权力和威望的正相关关系呢?

伯杰和他的同事发展了地位特征理论——这个理论我们将在第4章中讨论——作为上面问题的答案。我们在这里提及它,是为了说明伯杰等人(Berger et al. ,1977)提出的第三类问题如何联系到第一和第二类问题,就是贝尔斯及其助手在多年对基本小组演化进程进行观察后一直在问的问题。我们在图2.1中展示了问题类型和研究进程中的其他要素的联系。

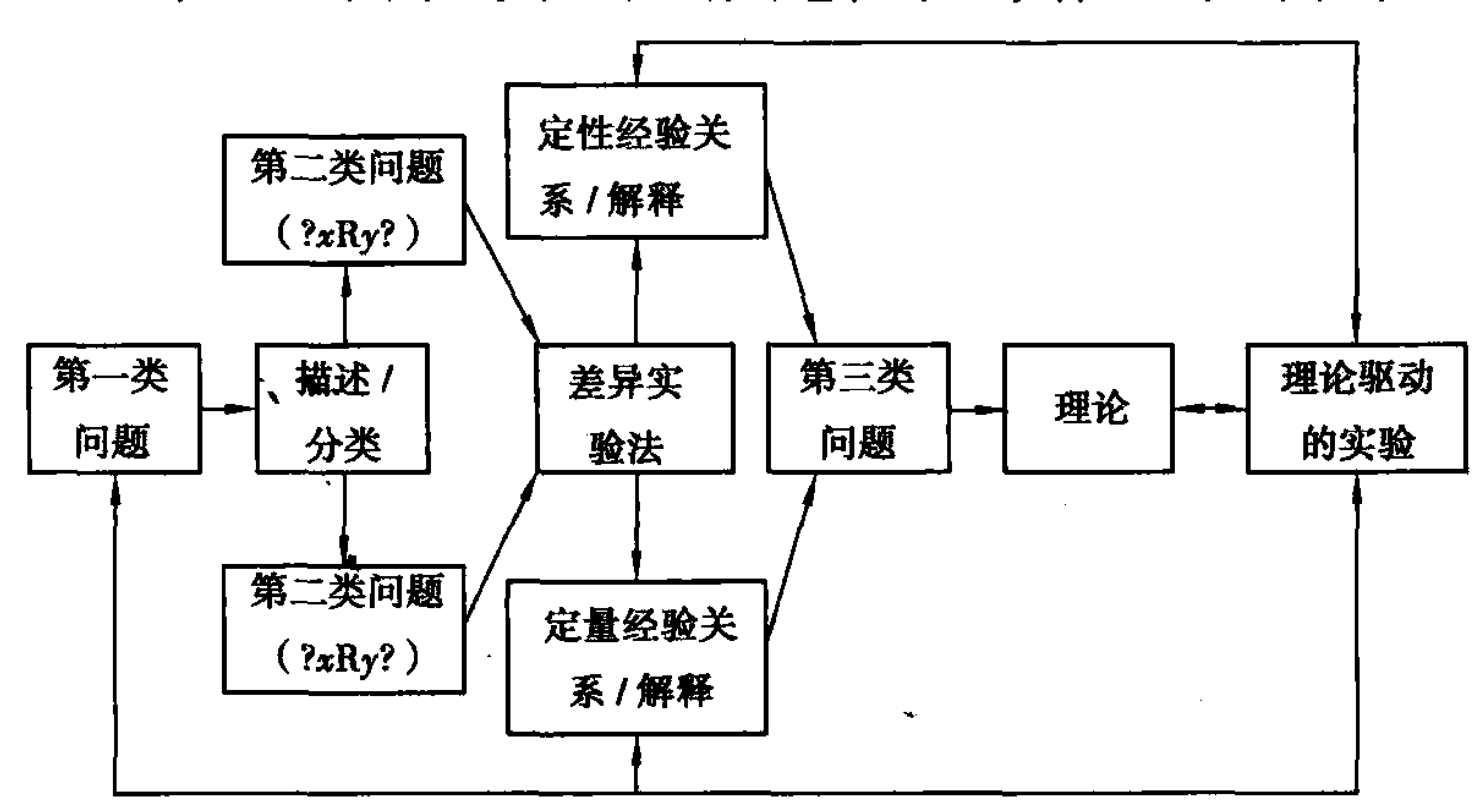


图2.1 研究过程中的问题、答案及实验

伯杰、科恩和泽尔底切(Berger, Cohen and Zelditch, 1966, 1972; Berger, Fisek, Norman and Zelditch, 1977)构建了一个理论性的研究项目,从而解释了一个在哈佛开始逐渐积累的研究项目中被分离和确立的现象。这个新理论必须被验证,并且通过解释原始理论未能解释的新现象而得到成功检验,进而提出新的问题。为了回答这些问题,更加新颖而复杂的理论产生了,当其被验证,并发现更进一步的现象时,进而再提出问题。这些相关理论和验证它们的实验都是理论性研究项目的具体细节,这些项目至今仍然活跃——在伯杰完成作为博士论文的第一项研究的差不多50年后。(Berger, Fisek and Norman, 1998;

Berger and Fisek,2006)。

熟悉科学史的读者知道科学是建立在前人工作的基础上的。必须如此,如果不这样的话,同样的理论和解释就会不断地被重复发现。尽管如此,构建科学却可以采取不同的形式。例如,不是所有的理论性研究项目都像地位特征理论一样,建立在先前实验工作的基础上。"基元理论"(Willer and Anderson,1981),一个关于社会结构和权力的理论,就历史性地建立在马克思和韦伯的理论基础上。"基元理论"中的某些部分与马克思和韦伯的部分理论相一致。

然而,即使建立在先前工作的基础上,也并不需要与其保持一致。正像在物理科学中反复看到的那样,一些显著的进步是从那些最尖锐的争论中发展而来的。在第4章中,我们回顾伽利略自由落体的研究,来说明它与社会科学实验之间的类似性。重要的是,伽利略的研究变革了物理学,因为指导其研究的理论出发点与亚里士多德的自由落体理论格格不入。伽利略确实是通过反对亚里士多德而建立理论的,而牛顿则通过赞成伽利略而建立理论。

到目前为止,我们已经考察了引发社会科学研究的各种问题,同时也探索了第一类和第二类问题如何构建成激发理论发展的第三类问题。我们已经描述了理论是什么,理论的抽象性如何与客观世界相联系,从而得到验证。实验为理论提供了最好的检验,但不是所有的实验都是为了检验理论而设计的。事实上,第1章中描述的三个实验没有一个是被设计用于验证理论的。它们都是经验驱动的实验,接下来我们将深入讨论这类实验。

3 经验驱动型实验

本章是探索实验设计逻辑的两章中的第一章，这部分将分析经验驱动型实验。这种类型的实验并不验证理论，而是由第一、第二类问题所推动，将新现象的发现和现象间关系的寻找作为目标。毫无疑问，我们将表明经验驱动型实验是发现现象的有力工具。

我们对经验驱动型实验的研究，源自于米尔（Mill[1843]，1967）的证据准则——经验主义实验方法论的逻辑基础[1]。我们将介绍米尔的准则，但只聚焦其中之一（差异法）作为方法示例。我们用米尔的科学方法构建识别问题，并讨论用费希尔（Fisher，1935，1956）的推断统计修改这一问题。我们也将解释费希尔对米尔方法精炼的必要性，并且确立好的经验主义实验设计的标准。最后，我们会批判性地分析 4 个范例的实验设计。分析的一个目的是表明经验驱动型实验[2]的作用和适用范围；第二个目的是明确优秀设计的特点，揭示出优秀设计与糟糕设计间的差异；第三个目的是解释如何将设计拙劣或执行不好的实验转变成设计更好、更具执行性的实验。

1　经验主义和经验主义者常被用作贬义词。我们以描述的方式使用它们。经验主义者获取知识的方法是"重视观察的重要性，重视通过大量观察和从这些观察中推广从而创造知识的重要性"（Cohen，1989：16；Popper，1962：21ff）。

2　虽然下面我们将关注米尔的差异法，但是所有这 5 个准则的逻辑与亨普尔（Hempel，1966）对发现法的描述一致。

米尔的准则和差异法

在《逻辑体系》(Mill [1843],1967)一书中,米尔以充分的推理提出了5条准则,用于发现本质规律。书中系统介绍了米尔的实验逻辑,描述了他的发现方法——求同法(methool of agreement)与差异法(method of difference)。

米尔的**求同法**用于确定逻辑学家所称的"必要性"。假设 A,B,C,D 和 E 是潜在的原因,并且 X 在以下各个情境中发生:①$\{A,B,C,D,E\}$;②$\{A,C,D,E,F\}$;③$\{A,F,G,H,I\}$;④$\{A,B,G,H,I\}$。A 是唯一四种情境都符合的潜在原因,B 不在情境2和3中,情境3和4中则没有 C,等等。因此,A 是 X 发生的必要条件,必要性以逻辑符号方式表示为 $X \rightarrow A$,读作"如果 X,则 A",但是依据经验,"X 仅发生在 A 存在的情境中",或者是"当 X 发生,A 则先于 X 发生"[3]。

米尔的第二种方法——**差异法**——是一种实验方法。与求同法相比,米尔将其看成更有效的发现工具。他陈述了如下准则(规则):

> 如果研究中的现象在一种情况下发生,而在另一种情况下没有发生,除某个条件外两种情况的其他条件均相同,这个条件只发生在前面一种情况;两种情况下唯一的差异条件就是影响因素,或者原因、或者原因(现象)不可缺少的一部分。(Mill [1843],1967:256)

差异法意在确定"充分性"。设想一个确立两种情况的实验:第一个包含条件$\{A,B,C,D,E\}$,第二个由条件$\{B,C,D,E\}$组成。在第一个条件(而不是第二个条件)后,观察到了 X。X 随着 A 的出现而出现,但 X 不出现,A 也不出现。这个观察证明了命题 $A \rightarrow X$ 是成立的:"如果 A,则 X"。依据经验,可以将

---

3 我们采用第2章中介绍的符号,A 和 X 是事件类。用更长的形式可将其读为,"当事件类 X 中的一个事例发生,那么事件类 A 中的一个事例发生在它之前。"为了便于呈现,我们使用更简短的形式。

其读为"当 A 出现,X 将跟随出现",A 是 X 的充分条件。

实验控制允许研究者确定现象的"因果顺序"(例如,确定 X 是原因还是结果)。在我们的例子中,研究者推断出 A 是 X 发生的一个原因,因为 A 是各条件间唯一不同的因子,并先于 X 出现。现在,A 通常被称为自变量,而 X 是因变量。在这里及其他地方,自变量被认为发生在因变量之前,并引发因变量的产生。如同我们的例子,差异法实验通常都至少有一个控制组(control group),还有一个或多个**实验组**(experimental group)与之相比较。

米尔明确了其他三个准则:**求同和求异并用法**(combined method of agreement and difference),**共变法**(method of concomitant variation),**剩余法**(method of residues)。求同和求异并用法只是将刚才讨论的求同法和求异法两者结合起来使用。设想两个或多个包含 A 因子的情况($\{A,B,C\}$ 和 $\{A,D,E\}$)随 X 发生。并且设想两个或多个不包括 A 因子的情况($\{B,C\}$ 和 $\{D,E\}$),没有随 X 发生。求同和求异并用法推论,A 是 X 发生的原因。

共变法比较 A 的变化值和 X 的变化值。如果 X 的值始终随着 A 的值变化而变化,那么 A 是 X 发生的原因。最后,剩余法用于多种原因或结果出现的时候。剩余是指其他所有的原因都已经确定而留下的部分。例如,假设研究者观测到 A,B 和 C 条件下的结果 X,Y 及 Z,但是在三个条件都不存在的情况下没有观测到这些结果。如果已知 A 和 B 分别是 X 和 Y 的原因,那么,剩余法将推论,Z 是 C 的结果。

我们再现米尔的推理,因为它对科学观念至关重要(Mill [1843],1967)。对米尔而言,实验研究是通过从表面上看起来混乱不堪的有序世界里**归纳**出规律,并发现它们(Mill[1843],1967:248)。归纳出规律是一种将世界**分解**成更小的部分的**分析**方法(从字面意义来看)。为什么能发现规律呢?这是因为经验现实是由现象间关系的有规律模式所组成的,但是在实验室之外,很多规律混淆在一起,形成了我们观察到的混乱的现实。有很多规律**潜藏**在现实里。但是,除非规律被挑明并辨识出来,否则世界上的现象将无规律地呈现在我们心中(Mill

[1843],1967:206)。

米尔体系的意义在于:实验将可观察世界的一般规律(那些不能被直接观察到的规律)分解成准确而细小的可观察到的规律。米尔关于这些规律的陈述是科学定律。因此,实验法发现了科学定律。

对于米尔而言,$A \to X$ 的规律不是由一次观察(X 出现在 $\{A,B,C,D,E\}$ 之后,或者不出现在 $\{B,C,D,E\}$ 之后)而确立起来的。**一个经验规律确切的是:在完全相同的条件下以完全相同的方式反复发生的一系列事件**。因此,经验规律的证据是通过重复得以证实的——通过反复实施完全相同的实验。关键是,米尔方法的缺陷就在于不能完全重验。

米尔的方法未能发现规律,其中有三个原因:第一,他的证明准则要求理想的规律。为了发现 A 是 X 的原因,X 必须总是发生在 A 后,当 A 没有被观察到时,则不会出现。事实上,米尔的方法并没有发现一个理想的规律。这里的问题并不在于社会科学研究不严密。**任何科学的研究都不曾满足米尔理想规律的要求**。考虑到测量误差,很显然理想规律不可能被观察到。

第二,米尔的差异标准不能得到满足,那无法遵守差异法描述中这样的命题“所有条件但有一个例外”。没有实际的解决方案能满足这个要求。其含义可以从下面的例子中看出来。米尔要求读者设想两种情境:

$ABCDE \to x$

和

$BCDE \to \sim x$

第一个情境被读为“当 $ABCDE$ 出现时,可观察到 x——x 的一种测量表征或出现标志”,第二个情境是“当 $BCDE$ 出现时,不能观察到 x”。因为在第二种情况下仅缺少 A,他假设例子的两种情境中“所有的条件相同但一个条件例外”。

然而,实验者的任务是在**真实世界**中,创造两种**经验**情境:已知它们完全相同,除了一个且仅有一个条件不同。我们假定一位实验者试图这样做,并且实现了上述第一种情境,然后实现第二种情境。但是因为两者实现的时间是不同的,所以有两

个不同之处。一般来说,任何两个经验情境彼此都不相同,并不像米尔假定的例子中仅有一个方面不同,而是有很多很多方面不同,其中任何一个都可能被称为原因,或者说现象背后原因的不可或缺的一部分。由此可知,米尔的方法不是一种切实可行的方法。

该方法面临的第三个问题是,研究者无法将观察结果作为规律。观察结果总是在时间和地点上是具体的,并且采用具体的术语加以描述。另一方面,规律仅包括理论概念的一般陈述(Willer and Webster,1970)。它们是一般观点间关系的表述,不涉及可观察现象间的具体关系。没有逻辑体系允许研究者从一份或上千份观察命题中归纳(或总结)出规律性的表述。事实上,这也是为什么包含规律的理论必须经过验证。一个科学家必须确定理论描述的关系能否可靠地重现。

读者应以怀疑的态度面对寻求规律的米尔体系。没有分析,这个世界是混乱的,但是有什么证据能表明存在潜在的规律并且这些规律共同造成了混乱呢?一个也没有。然而,米尔的差异法的逻辑要求理想的规律,那些追随米尔的人时常害怕了解到:在 A,B,C,D,E 出现后,X 未能出现,以及在仅有 B,C,D,E 的条件下 X 有时也会出现。因为差异法实验产生的结果与米尔差异法准则并不一致,那些采纳他的关于科学观点的人不得不寻求不那么苛刻的规则。

费希尔和差异可能性

罗纳德·A. 费希尔(Fisher,1935,1956)用概率理论和推断统计把米尔的差异法从不可能发现理想规律的沼泽中拯救出来。事实上,米尔已经表明他的方法能发现不那么完美的或可能形式的规律:A 可能是 X 的原因。但是,可能性规律的重要性仅在于"向普遍的真理逐渐迈进"(Mill [1843],1967:387),相比较而言,费希尔的目标是寻找可能的规律。

费希尔对实验设计作出了重要贡献,包括对控制组和实验组的随机分配。对于费希尔的农学研究而言,随机分配意味着

将某一或另一品系的小麦随机分配到某一或另一块地。在社会科学里，随机分配是指将人随机分配到实验组和控制组。不管哪种情况，随机分配都成为了采用实验数据的统计检验的逻辑基础。统计检验评估了自变量和因变量之间是否有联系，而随机分配至少排除了因变量发生的其他原因。

让我们再现费希尔的推理。实验组和控制组的随机分配允许这样的推论，任何特定测量的平均数和方差在**分配点**上的两组都是相似的。随着随机分配，分配总体的成员以相同的可能性被分派到任意一组中。因此，两组的平均数和方差应该几乎相等（它们也应该与总体的平均数和方差相等，即使两者一般是未知的）。依据大数法则，相似性应随着小组规模的增大而增加。

假设在2006年春天，一个由22 400名亚利桑那大学全日制本科生组成的潜在被试的群体（或者说总体），其中，53%是女性，28%是少数民族。现在，由研究者随机抽取群体中的200名成员，并且随机分配他们到控制组和实验组。那么每个小组100名成员中将有53名（53%）女性和28名（28%）少数民族，且其他的百分比特征也相类似。此外，如果小组规模大小从100人增加到200人，那么两个组中女性人数、少数民族人数也会随之增加。

在差异法实验中实验组和控制组一般是由称为**处置**（treatment）的介入方式加以区分的。这种处置实施在实验组，而不是控制组。例如，二战期间所进行的**美国士兵**的研究（Stouffer et al.，1949），士兵们被随机分配成两组。实验组观看电影《为什么我们要战斗》，而控制组没有观看这部电影。电影的内容是有关法西斯主义的罪恶。问卷项目的平均反馈比较，显示影片取得了意料中的效果。实验组的士兵比控制组的表现出更强烈的与德国战斗的意愿。因为士兵是被随机分配到实验组和控制组的，可以推测，如果没有看过影片，那么士兵们在抗击德国的意愿上没有差异。

费希尔对差异实验的再设计引入了发现逻辑的重大转变。从上面的描述来看，随机分配被试到实验组和控制组，这一实践试图完成样本平均数相等的假设。检验的**关键**假设就是样

本平均数**没有差异**,也称之为**零假设**(null hypothesis,也译为虚无假设)。更精确地说,零假设认为,在缺乏处置作用的情况下,如果观察到样本组之间的差异,那么它们是由随机分配产生的偶然差异所导致。

费希尔重新设计了米尔的差异实验。第一,他取消了米尔实验寻找绝对规律的要求;其次,他用更弱的似然规则或概率规律替代了米尔的准则。为了将更弱的规则运用到实践中,费希尔不得不克服一个问题:如何确定似然规律?概率论似乎提供了解决方案。

从概率论的观点来看,在任何实验里被分配到实验组和控制组的被试被看成两种分配,这两种分配来源于所有可能分配的(假设)总体中。对于总体,存在一个(假设的)平均数 μ 和方差 σ。如果研究者已知平均数和方差,他们可以设计一个统计检验,以确定实验组和控制组间由处置所引发平均数差异的可能性。但是方差不是已知的,也从未能得知,因为它是假设的,而非真实的分配总体的差异。概率理论解决问题的尝试似乎走入了死胡同。

威廉·戈塞特在吉尼斯酿酒厂工作,是一位有进取心的统计员,他设计了一套解决方案,用样本方差替代未知的、也不可能知道的总体方差。戈塞特的高明之处在于运用样本数据来计算每一组的标准差,这种计算对于很多人来说是熟悉的。分配总体的标准误差的估计,即**标准误**(standard error),是由计算样本大小和权重标准误差得来的。于是标准误通常用于计算统计量 t。t 值是实验组的平均数减去控制组的平均数除以标准误差的加权平均数[4]。对于已知大小的样本,戈塞特的统计量 t 可用于确定平均差异偶然发生的可能性。回到前面,在实验处置之前,实验组和控制组的平均数被假定是相等的(或者几乎相等)。因此,可以利用“不可能偶然发生”的 t 值(例如 100 次中少于 5

4 n 值标准误的加权平均数和 t 值以下列方式计算:$\sigma_{(x_1-x_2)} = \sqrt{\dfrac{N_1 s_1^2 + N_2 s_2^2}{N_1 + N_2 - 2}}$

$$t = \frac{X_1 - X_2}{\sigma_{(x_1-x_2)}}$$

次)来推断实验处置是否是结果的一个可能原因[5]。

费希尔修正了米尔的差异法,摒弃了其绝对规律的标准,取而代之的新标准是:**如果不能归因于偶然性,那么观察到的差异就是一个可能规律的证据**。纯粹主观臆断的值 0.05 和 0.01已经成为社会科学中不可能性的标准(Leahey,2005)。例如美国士兵的实验,费希尔仅仅要求实验组想要与德国作战的意愿比控制组更显著。相对而言,米尔则要求实验组的所有人都想要抗击德国,而控制组没有一个人受到鼓动。但是,米尔要求的结果是不可能找到的。因此,与米尔方法不同,费希尔的方法是切实可行的。事实上,当 n 足够大时,样本平均数彼此间可能非常接近,且结果可能大量交迭。然而,t 检验仍然可能表明自变量与因变量的关系。

尽管费希尔的标准已经取代了米尔的准则,然而经验驱动型实验的目标显然没有发生变化。主要目标是通过差异法的运用发现规律,形成科学。然而,现在规律不是如同米尔要求的那样绝对与确定,而是局部的,不太可能归因于偶然性的。

尽管米尔和费希尔之间的有些目标已经发生了变化。发现规律的目标,连同在所发现的规律之外再现现实的希望,已经迷失了。它们的迷失表明,研究者不能用差异法实验从观察报告中发展科学。我们现在将要表明的是,科学发展不是经验驱动型中差异法实验的正确目标。

经验驱动型实验的标准

正如费希尔所提炼的那样,经验驱动型实验已经被证明是非常重要的科学工具,因为它是一种寻获关系规律的非常有效

---

5 当代的啤酒饮用者感到惊奇的是,他们最喜欢的酿造啤酒之所以在纽约或波卡特罗品尝起来是一样的,要归功于戈塞特。吉尼斯酿酒厂关注于开发更好的方法以比较啤酒样品,从而更好地评价和控制啤酒质量。为了确保其竞争者不知道戈塞特更好地评价样品的秘密,吉尼斯酿酒厂禁止戈塞特公开发表他的方法。鉴于此,又由于他熟悉同时代的顶尖统计学家,并且能经常与他们交流,戈塞特便以笔名“学生”发表了他的工作。许多统计学文献把戈塞特的方法称为“学生的 t 检验”。

的方法。当然,对于特定研究来说,它的有效性依赖于其是否被恰当地运用。在这部分,我们要确立经验驱动型实验有效性的标准。如果这些标准无法满足,结果和解释将会受到人为因素的影响。人为因素是由研究环境的特点、实验者或者实验中的被试所造成的系统影响,它可能导致研究者推测观察结果的原因时犯错误(Kruglanski,1975:103)。他们可能错误地将实验结果归因于受实验控制的变量,或者失败地将结果归因于不受控制的变量——或者两者兼有。这是**内部效度**的问题:实验是否已经排除了现象的其他原因,以至于结果可以恰当地归因于受实验控制的变量[6]。

正确地设计差异法实验的标准是直接遵循米尔和费希尔的实验逻辑。我们用四条准则加以说明:

1. 创建至少两种尽可能相同的初始研究条件。
2. 引入两种条件间的单一差异,然后观察结果。
3. 将有关结果的所有推论限制到单一差异的效果上。
4. 仅当其不可能归因于偶然性时,推断关系规律。

当实验涉及人时,前两条准则能以两种非常不同的方式得到满足。第一,最常用的步骤就是随机将人分配到实验组和控制组,以至于除了分配过程中发生的随机差异外,被试在一开始都是相同的。美国士兵的实验就采取了这种方式。其次,引入研究者感兴趣的差异作为实验处置。影片《为什么我们要战斗》是美国士兵研究中最关键的差异。实验组的被试观看了影片,而控制组的士兵没有看。在本章随后讨论的阿施(Asch,1958)实验中,其更精细的设计拥有多个实验组。

也有第二种方式,能使第一条和第二条准则都得到满足。从已知两个不同的小组开始,在相同的条件下研究它们。按照这种方式,研究者通过研究两种(或者多种)不同类别的人,引入差异。例如,后面讨论的辛普森(Simpson,2003)实验中,所有的被试参与完全同样的实验博弈。不同的是,其中一组为男性,而其他组是女性。

---

6 第5章讨论了人为因素,人为因素的来源(包括需求特征、被试偏差和实验者偏差),以及减少和消除人为因素的技术。

满足第三条准则就更直接了，但是，很多实验却未能满足它。例如，在接下来讨论的一系列的实验中，米尔格拉姆(Milgram，1965，1974)主张，实验者的合理需求能解释被试对同伴实施严厉电击的意愿。但是米尔格拉姆的实验没有系统地设置合理性变量，并且他没有设置控制组，合理性是缺失的。最后，满足第四条准则：关系规律不可能归因于偶然性，就像前三条，这一条仅能在设计很好的实验中得到满足。现在我们应用这四条准则，分析四个差异法实验的设计。

阿施的从众研究

所罗门·阿施关于从众和判断失真的研究是一个受到广泛认可的经典实验。阿施被第二类问题激发，他的研究目标是发现社会条件(因素)，这些条件会“促使人们抗拒或屈服群体压力，在后者被知觉到**违背事实**的时候”(Asch，1958:174)。阿施在有关法西斯主义和第二次世界大战后果的著作里断言，人们之所以屈服群体压力是因为“显而易见的社会结果”(Asch，1958:174)。直到今天人们仍然研究他的实验。我们将评价在他的设计中违背好的设计准则的关键要素。

阿施最早的研究是八人坐成一排(或半圆形)，但是其中七人是实验者的同伴。这些人面对着实验中使用的刺激材料坐着。刺激材料是研究助手按顺序举起的18对卡片[7]，每一对卡片的第一张有一根称为**标准线**的垂直线，第二张卡片有三条垂直的对照线。一条对照线总是与标准线相同，另外两条线明显短于或长于标准线。大家的任务是判断哪条对照线与标准线一样长，并向研究人员报告他们的判断。

阿施的设计是经典的差异实验。实验处置是要求参与者按顺序大声地汇报出他们的判断。唯一的被试坐在第七或第八个位置上。到18对刺激卡片中的第12对时，那些实验者的同伴都

7 标准的阿施任务已经衍生出许多变化。有些采用幻灯片把线条投射到屏幕上，而其他一些则将刺激材料投放到计算机上。

给出同样的错误答案,然后要求被试大声地给出自己的答案。已给出的大多数一致的"错误"是为了制造社会压力,即自变量。阿施将被试的反应,不管是正确还是错误,视为因变量。

阿施的控制条件是让被试和7个同伴写出他们的判断(也就是说他们的判断是保密的)。另外,实验和控制条件是相似的。实验设计允许阿施测量控制组的错误率,并与实验组相对比。两者间的差异就能够测量社会压力的效果。

面对一致的多数人,50名被试共有32%的错误率,而且所有错误都与多数人的反应方向相一致。在控制组中37名被试的错误率少于1%(大约0.68%),这一数字要比实验组低31个百分点。实验组的被试平均每人有3.84个错误,而控制组被试平均每人仅有0.08个错误!

阿施研究的从众,发生在人们改变自己的判断,盲从其他多数人违背事实的判断时。理解阿施实验产生的效果,以及研究设计如何准确地阐释这种效果,这一点是很重要的。阿施关注的不是被试能否作出正确的判断。事实上,所选线条的长度明显不同,因此控制组产生的错误判断可忽略不计。相反,他们关注的是在多数人一致说出明显违背事实的判断时,被试的判断是什么。

这个设计是很优秀的,因为它明确地从"影响"中分离出了"权力"——两个通常一起发生的现象,并且通常相互混杂在一起。尽管阿施在一些论文题目中使用术语"影响","影响"仅出现在信念发生变化的时候(比较Zelditch,1992)。阿施对被试的事后访谈表明他们的信念没有受到影响——几乎没有被试真的相信他们说出的错误判断是正确的。

阿施对抗拒群体压力与迎合群体压力同样感兴趣。为了研究抗拒,他设计了一组有不同自变量的对照实验组。阿施分别采用1,2,3,4,8或16个同伴,以改变实验者同伴的数量。然而,一个同伴几乎没有效果,最好的效果发生在3个和4个同伴的情况下,而随着人数的增加效果略微减弱。应用米尔的共变法准则,从众行为随着产生分歧人数的变化而变化,其关系不是线性的,而是曲线的。

阿施也研究了不一致的多数人,他采用了两种方式。第二

个实验被试由一个同伴所替代，或者一个同伴经过指导能作出正确的判断。一致多数的情况下错误率超过30%，而对于两个不一致多数的设计，其错误率分别降至10.4%和5.5%。最后，阿施取消了开始的实验设计，在16个被试的小组中安排了唯一一个犯错误的同伴。这个同伴完全没有影响被试的判断，反而成了其嘲笑的目标。

阿施的实验设计有几个优点。他开发了自变量和因变量，以至于自变量（群体压力）能发生变化，也使得因变量（错误率）的变化可以很容易被测量出来。阿施研究的自然要素满足了差异法逻辑的第一和第二条准则，因为控制组和实验组仅有一个因素不相同，即判断是以保密的还是公开的方式被报告。应用第三条准则，这一特征加强了联系的依据——在同伴的判断和被试的判断报告之间的联系。然而，阿施没有提到随机分配，仅仅报告了差异。他没有用检验统计来评估差异源于偶然的可能性。

其他两个优秀设计的优点也应该被提到。第一，像阿施这类设计通过"强力战术（sledgehammer approach）"产生了清晰的结果。在强力战术中，自变量的效果非常大，以至于它掩盖了由于取样和测量误差导致的因变量的变异。阿施的设计产生了强大的效应，结果，很少有人质疑是否是群体压力影响判断——尽管没有随机分配。第二，实验设计有着很高的**现实性**，因为被试是处于一个现实而有效的实验条件中。有时实验现实要与"世俗现实"相对比，这意味着实验应类似于实验室之外的特定条件。

阿施实验在今天依然重要，至少有一部分是重要的，因为许多设计要素是合理的，并且实验产生了很强的效果。但是依据现在的标准和常规，他的设计有多么优秀呢？今天，要求随机分配和更加精确地呈现定量数据。例如，据他的报告称，控制组的错误率是第一个实验组汇报结果的2%。这个差异是非常大的，效果也非常明显。但是目前标准要求的，不是这种对比，而是统计检验，这种检验能揭示出各实验组间被观察到的差异不可能归因于偶然因素。在比较人数分别为3，4和5的小组，并且实验组间的差异不是很明显的情况下，统计检验显

得尤为重要。期刊编辑也要求更进一步的实验来明确组的大小与从众程度之间确切的曲线关系和形状——这也是我们乐于知道的事情。幸运的是,阿施做出了强有力的实验效果,精心刻画了实验组与控制组之间的差异,这些没有给人留下任何疑问,至少对我们而言,他测量的效应与群体压力造成的从众是一致的。

米尔格拉姆的服从实验

在这一部分我们将考察斯坦利·米尔格拉姆(Milgram,1974)关于服从权威的实验。大学心理学教科书总会描述米尔格拉姆的研究,有些高中心理学课程也会讨论这个话题。和阿施一样,米尔格拉姆的研究也是由第二类问题所激发:什么条件影响对权威的服从(Milgram,1965)? 和阿施一样,在希特勒第三帝国残暴统治期间所发生的问题激发了米尔格拉姆的研究兴趣。世界上很多人在质问,有几百万人在纳粹死亡集中营遭受凌辱、折磨和毒气伤害,大屠杀是很多邪恶的人所为,还是许多普通人屈服于少数邪恶权威的结果呢? 米尔格拉姆认真地研究了这一问题,设计实验来发现权威的局限。

不像许多在大学校园里实施实验的那些人,米尔格拉姆是在校外场所实施一些研究的。他的研究程序也不同于大多数研究者所采用的程序,而是从一般的普通民众中招募被试,包括工厂工人、体力劳动者、商人、专家和其他人。学生被特别排除在外(Milgram,1974:15)。此外,正如我们所知,米尔格拉姆的研究是一系列跨国研究中第一个研究权威和面对权威的反应如何发生变化的。德国的研究紧随美国的研究之后,米尔格拉姆认为在德国服从程度更高。米尔格拉姆从未在德国重复过他的研究,但是曼特尔(Mantell,1971)从在德国所作的服从研究中发现服从率高达85%。

最著名的米尔格拉姆实验(实验2)是这样设计的(Milgram,1974:13ff):两个获得酬劳的志愿者进入实验室,其中只有一个人是被试,而另一个人是实验者的同伴,而且被试

并不知道这一情况。实验者向他们解释说“记忆与学习”的研究涉及两个角色：“老师”与“学习者”。任务要求是：老师面对学习者读一系列的单词对，并要求学习者在看到每对单词的第一个词后，回忆其中的第二单词，学习者出错后，老师按照指示电击学习者，学习者接下来再出错，老师则加重电击强度。任务被解释完毕后，假装随机抽签，将实验者的同伴安排成学习者角色，将被试安排成老师角色。

被试看到实验者将学习者绑在椅子上，或者像有些实验，帮助实验者将学习者捆绑好，之后随同实验者到另一房间里。被试坐在“电击发生器”旁，控制面板上有30个开关，控制着30个水平的电击，从15～450伏以15伏的幅度递增。电击发生器上的数字附以标签，从“轻微电击”增大到“危险：强烈电击”。最后两级（435和450伏）标以“XXX”。

在米尔格拉姆的任何实验里都没有人真正遭受过电击，但是从75伏（第五级）开始，实验者的同伴（学习者）被要求按一套言语模式，表达出随着电压增加越来越不舒服的感受。到150伏时，同伴要求离开；到300伏时，他开始尖叫；当电击发生器达到最大伏特时，他乞求被试停下来，并且拒绝回答接下来的问题（Milgram，1974：56－57）。因变量是被试在拒绝继续执行前实施的最大（标示）的电击级数。

实验者在与被试的互动中扮演了一个重要而积极的角色。当被试对继续执行更高电压级表现出迟疑时，实验者坚定地说出下面的激励话语：

激励1：“请继续”或者“请继续下去。”

激励2：“实验要求你继续。”

激励3：“你绝对有必要继续下去。”

激励4：“你没得选择，必须继续下去。”（Milgram，1974：21）

如果被试说这对学习者有伤害，实验者则说：“尽管电击可能非常疼，但是不会产生永久的组织损伤，所以请继续”（Milgram，1974：21）。实验者根据需要逐一说出激励语1—4，以此来促使被试增加电击强度。如果被试辩解说学习者不想继续下去，实验者则说：“不管学习者愿不愿意，你都必须继续下去，直到他正确地学会所有的单词对，所以请继续”。然后按

需要,补充激励语1—4来促使被试增加电压。

米尔格拉姆采用不同于基本实验中的18个实验,确定其中之一(实验11)作为唯一的"控制"(米尔格拉姆的原话)条件。在控制实验中,允许被试选择电击的级数。米尔格拉姆没有报告实验中是否使用了激励,但是既然被试可以选择电击级别,激励就与之没有什么关系了。

受控条件下,被试执行的平均最大电击是82.5伏。有23个被试(57.5%)从未达到过75伏的级别,而在这一级别实验者的同伴开始了言语反应(Milgram,1974:72);12个被试(30%)使用了高于75伏的电击级别,而仅仅1个被试(40个当中)达到了最大的电击级别。相反,上面描述的实验2中被试执行的平均电击级别为367.9伏,40个被试中有25个(62.5%)达到了最大电击450伏。事实上,控制组的值远低于接受了其他设计的组。

毫无疑问,米尔格拉姆的实验引发了服从,但是他的实验设计是差的。被试没有被随机分配到实验中,自变量没有被正确地测量,贯穿这个实验的变量中不只一个发生了改变,而且米尔格拉姆从来没有报告统计检验。事实上,这一系列的实验没有满足上述任何一条准则。例如在他的控制实验和实验2之间至少存在两个差异。在控制条件下,被试自己控制电击。相对而言,实验2中实验者详细说明了电击等级,而且,当被试犹豫的时候,要求其继续执行电击。

贯穿所有的19个实验,在设计上很少、甚至没有将服从与权威联系起来。权威一般被定义为"合法的权力(legitimate power)",但米尔格拉姆却没提供给读者权威的定义,并且在实验过程中没有使用术语"合法的权威(legitimate authority)"。可以肯定的是,实验者穿着实验服,以权威的方式行事,但被试从实验者的行为和耶鲁大学教授(如果已知)的身份中能推论出什么,这一点是未知的,也没有被报告。虽然有实验后访谈,但是因为没有把准确的内涵联系到合法的权威上,所以有一点是不清楚的,米尔格拉姆如何确定被试的反应针对的是实验者合法性还是强制性权力。

此外,因为权威是一个自变量,所以这些实验明显地违反

了上述的第二条准则。正如上面所提到的，实验 11 被称为“控制条件”，但是控制变量不是实验者作为合法权威的地位。米尔格拉姆创建了错误的控制条件。因为他的目的是表明人们服从权威的行动，所以控制组应该有一个实验者所提的不合理的要求，但是研究从未实施过那个条件。

米尔格拉姆声称实验者是合法的权威形象，但是他不经意提到的一些事则表明其实验中缺乏实验者的合法性。实验不是在耶鲁大学里，而是在康涅狄格州的桥港附近实施的，在那里实验者声称代表完全虚拟的“桥港研究协会”。在这些研究中两个被试的评论也质疑了合法性。第一个人写道：

> 我可以退出这个该死的测验吗？他可能昏倒了！我们太愚蠢了，没有核对这个协议。我们怎么知道这些家伙是**合法的**。没有家具，空房子，没有电话。(Milgram, 1974:69，额外加入的强调)

第二个人表明：

> 我怀疑我自己得出的判断(关于即将到来的)。我已经质疑操作的合法性和参与的后果。(Milgram, 1974:69，额外加入的强调)

重要的是，米尔格拉姆发现了在桥港和在耶鲁大学相似的服从水平。

米尔格拉姆的设计还有其他的问题。就像阿施的研究，米尔格拉姆的实验也对因变量有强大的影响，这是很大的优势。然而，与阿施的研究不同的是，实验设计时没有明确自变量。效果是由合法权威造成的吗？或者米尔格拉姆的被试产生了对强制权力的反应吗？

一旦实验开始，就有一件事是明确的：被试有充分的理由相信，实验者确实残酷无情。为了得到数据，实验者将一些人捆绑起来，并对他们施行危险的，甚至可能致命的电击。此外，面对犹豫，残酷无情的实验者坚称“你没得选择，必须继续下去！”(第四个激励)。然而，激励没能详细地说明不服从的后果，强制性的威胁也没有被明确地表述。**预期反应定律**(Law of Anticipated Reactions)(Naggel, 1975)表明，人们是通过预期到权力人物要求他们去做什么，以及如果他们未能遵从，权力

人物将要对他们做什么,从而遵从强制权力的。

与合法权威相比,强制权威更可能是服从的原因?或者米尔格拉姆是正确的,合法权威导致了服从[8]?又或者说还有第三种或者第四种原因?因为他们的实验设计得不好,所以这些问题无法从米尔格拉姆的研究中得到解答。

社会行为的自动化

这个部分分析了三个实验设计,巴夫、陈和伯罗(Bargh, Chen and Burrows,1996)曾使用这些实验,研究社会行为的**自动化**。若社会行为由当前某一情境里的物体或事件所激发,且没有通过有意识的知觉或判断过程,则社会行为是"自动的"(231)。换句话说,正在激发的物体或事件与行为之间的联系是直接的,不可识别的。实验的目标是揭示自动化——发现被试未能意识到的刺激是否影响了其行为。发现或发展更特别的现象是第一类问题,将刺激与行为联系起来是第二类问题。尽管巴夫等人广泛引用相关的研究和哲学家威廉·詹姆斯的评论,但是理论在实验设计和推动他们形成假设方面并没有发挥作用。

巴夫等人的三个实验共同具有下列设计要素:

1. 被试被随机分配到实验组和控制组。

2. 启动条件——自变量——隐藏在更大的任务中,它的完成就是实验者声称的实验目的。

3. 因变量的测量得到了伪装,并且在2/3的情况下测量的时机要么出现在实验的两个部分之间,要么出现在实验之后。

第一个实验研究礼貌和粗鲁能否被启动,以至于被试表现出一种或另一种行为,以及在被试没有意识到的情况下,礼貌或粗鲁的行为已经产生。被试被告知,实验被分为两部分,都

---

8 怀疑不同解释的可能性的人应该看一下米尔格拉姆实验的影片(通过宾夕法尼亚州立大学媒体销售网址 http://www.mediasales.psu.edu/. 可以看到这部影片)。

是关于语言能力的。实验有三个初始条件:礼貌的、中性的和粗鲁的。每一被试都将做一个由5个混乱排列的单词组成的"乱句"(例如:"他它隐藏发现立即地"),以及30个可选用的单词,要求每次只选用一个,构建30个由四个单词组成的句子。30个单词中,有15个分别代表3种不同的条件:粗鲁启动的单词包括**大胆的**、**粗鲁的**、**打扰**、**无耻的**等,中性的单词包括**练习**、**偶然地**、**熟练的**、**清晰**等,礼貌启动的单词包括**尊重**、**荣幸**、**承让**、**礼貌**等,这三组共计15个单词代表了自变量的变化。

因变量用以下的方法加以测量:告诉被试,完成测试之后,到走廊找实验者,然后第二部分的实验就开始了。实际上,实验者在半开的门旁做出一副显然是在向第二个被试详细地解释实验的样子(第一个被试看不到第二个被试)——第二个被试是实验者的同伴。实验者的解释会一直持续,直到被第一个被试打断,或者10分钟的时间已经到了。因变量是被试打断实验者之前的等待时间。

研究者用两种方式鉴别效果。第一种,他们比较了打断前时间的平均数(以秒为单位)。结果是这样的:粗鲁的=326,中间的=519,礼貌的=558 。方差分析表明效果显著,并且通过F检验表明在粗鲁处置上平均时间显著不同于其他两组。第二种测量——没有打断实验者的人数百分比,更进了一步,因为35个被试中20个(57%)没有打断主试,而是足足等了10分钟(也就是600秒)。礼貌组中少于20%的被试打断实验者,中性组中大约35%的被试打断实验者,而粗鲁组则超过60%。这些结果明显与假设的启动效应相一致。为了评估它是否是自动发生的,或者这些行为差异是否由实验者的行为解释作为中介,研究者询问被试实验者是否是有礼貌的。实验者报告在被试对实验者礼貌的评价上不存在组间差异。

第二个实验的设计是为了激活"老年人"的刻板印象,再一次采用"乱句测验"。这一次实验组得到了启动"老年"的单词,而控制组得到了中间年龄的单词。重要的是,对于这两个组都不包括涉及时间和速度的单词。因变量是被试在实验明确完成后,走向大厅所用的时间。以秒为单位测量,实验组平均为8.28,而控制组平均为7.30,差异显著性大于0.01 。其结

果的另一种解释就是心境的干扰。为了测试心境,一组新的被试采用相同的"乱句测验",然后进行情绪状态的测验。在情绪状态上没有发现显著性的差异。因为在实验组或控制组中都没有提到时间和速度,且心境也没有介入其中,可以得出测试的效应是自动化的。

第三个实验试图测量非裔美国人对比欧裔美国人的刻板印象效应。没有被试是非裔美国人。在进行冗长乏味的实验任务之前,一张面孔短暂地出现在被试的电脑屏幕上,如此短暂以至于被试不能回忆起看到过这张面孔。在被试进行实验一段时间后,电脑声称在那个时点的所有数据都已丢失,被试必须重新开始。两个编码者注视隐藏的摄像机,来评价被试面部的敌意反应,他们不知道实验条件。我们所说的"不知道实验条件"是指编码者不知道被验证的假设,也不知道显示了哪张面孔。接受非裔美国人面孔启动的被试具有更显著的敌意。然而,被试不能回想起看到了这张面孔,研究者也测量实验组和控制组被试的种族态度。因为他们在实验组和控制组之间没有发现种族态度上的差异,实验者得出这个效应是自动化的。

这些设计很好的实验很容易地满足了前面列出的四条准则。几个实验特征提升了实验的质量。第一,随机分配用于确立实验组和控制组一开始就相同的可能性,但是可能存在偶然差异。而且,随机分配满足了采用统计程序来评价观察到差异的前提条件。第二,实验组和控制组仅在启动条件上不同。然而,如果被试意识到了实验目的,研究者就因此认为任何观察到的差异都是对启动事件的自动化反应,这种说法则是错误的。因此,研究者将启动事件隐藏在更大的任务中——指导语中,给出一个表面上的故事来遮掩实验目的。因为存在**需求特征**(demand characteristics),被试会推断实验者的假设,并试图使自己得到满足,而这一步骤应该排除了可能发生的效应[9]。第三,他们通过每次测量因变量,使其看起来好像不是实验的

9 奥恩(Orne,1962)采用术语"需求特征"表示那样一些例子,即被试推断实验者的假设,担当"好被试"的角色,并且以趋于证实假设的方式行事。需求特征将在第五章讨论。

一部分，从而隐藏了因变量的测量。如果与自动化一样微妙的现象是研究目标，这一步骤毫无疑问是必要的。如果测量得到有效的掩饰，被试有目的做出（或未能做出）他们认为实验者期望或期待的行为，就不能帮助或干扰实验者。有些略微不同的说法，如果被试没有意识到被测量的行为，需求特征就不能影响测量。第四，实验者采用"双盲（double-blind）"程序，减少了实验者偏差的可能性，也就是说实验者已知或未知的行为造成被试行为的偏差，从而支持了假设。

最后，实验者不满足产生一些效应，并且希望读者可以从实验—控制组的设计中看出，效应是自动化的。反之，访谈可用于鉴别和排除一些似是而非的因素，这些因素可能在启动事件与随后的行为之间建立起有意识的，而非自动化的联系。

实验室里的性别、恐惧和贪婪

这部分的研究探索了男性和女性的动机是否不同，从而使他们对特定的社会困境问题做出了不同的反应。这一观点是第二类问题：因素 *A*（性别）与 *X*（对社会两难问题做出的反应）是否有关系？布伦特·辛普森（Simpson，2003）以下列方式采用差异法逻辑：所有被试经历的条件都是相同的，所有人都要玩三个相同的困境博弈，唯一不同的是被试是男性，或者女性。因此，实验设计满足了准则 1，2 。辛普森的设计与前三个实验设计唯一的不同之处是：他的设计是一个混合设计，并且结合了差异法（经验驱动型）和理论驱动型实验的特点。

简要回顾实验中使用的一些博弈理论的要素，将有助于表明理论是如何影响设计的。表 3.1 以"标准形式"呈现了三个博弈。每个博弈由两个人来参与，他们同时做出决策，并且不知道其他人的决策。在这个实验中，要求每个人通过选择标记着 C 和 D 的两个选项中的一个，来获得尽可能大的收益。C 代表合作，D 代表背叛。对于每个博弈，设定个体 1 会选择 C（上排）或者 D（底排），而个体 2 在 C（左列）和 D（右列）之间选择。

表 3.1 三种社会困境的支付矩阵

	囚徒困境		恐惧(无贪婪)		贪婪(无恐惧)	
	C	D	C	D	C	D
C	3,3	1,4	4,4	1,4	2,2	1,4
D	4,1	2,2	4,1	3,3	4,1	1,1

来源:辛普森(Simpson,2003)。

表 3.1 中的三个博弈是**战略决策情境**(strategic decision situation)。我们所说的"战略"的意思是每个人所获得的收益是由双方共同选择来决定的。先考虑囚徒困境博弈(以后称为 P/D)的收益。每一逗号左边是博弈者 1(以后称为 P_1)的收益,每一逗号右边是博弈者 2(以后称为 P_2)的收益。例如,当两者选择 C(合作),都得到 3,但是当 P_1 选择 C,P_2 选择 D(背叛)时,P_1 仅得到 1,而 P_2 得到 4。颠倒选择则收益相反:当 P_1 选择 D,P_2 选择 C 时,那么 P_1 和 P_2 分别得到 4 和 1。最后,如果他们相互背叛,两者均得到 2。表 3.1 中其他两个博弈也同样是策略决策情境,因为每个人的利益都是由双方共同选择来决定的。

被试怎样做才能从 P/D 博弈中获得最大的收益呢? 首先要问:"当**对方**合作时最好的选择是什么?"然后要问:"当**对方**背叛时,最好的选择是什么?"现在比较这两个问题的答案,答案是一样的吗? 如果是,那就是最好的策略。因为博弈是对称的,其中一个博弈者发现的答案就是两个博弈者要找的答案。

现在让我们了解实验步骤,从 P_1 的视角来解释步骤。如果 P_2 合作,P_1 该做什么? 当 P_2 合作的时候,P_1 通过合作将获得 3 分,或者通过背叛将获得 4 分。因此,P_1 应该选择背叛。如果 P_2 背叛,那 P_1 该选什么? P_1 通过合作将获得 1 分,或者通过背叛将获得 2 分。因此,也应该选择背叛。答案都是"背叛"。因此,不管 P_2 的决策是什么,P1 都应该选择背叛。因为这个博弈是对称的,P2 也应该选择背叛。一般而言,如果不管其他人选择哪个选项,有一个选项总是更好的,那么在这里,那个选项叫做**占优策略**(dominant strategy)。

现在,你就会理解为什么 P/D 博弈被称为两难推理。正如刚刚看到的,两个博弈者选择背叛是理性的,当他们这样做的时候,每个人得到 2 分。这个博弈是一个两难的选择,因为彼此合作的话,得到的不是 2 分而是 3 分。就像他们在辛普森实验里所做的那样,如果两者彼此只玩一次,那么两难的问题就不能得到解决[10]。

辛普森从两种非正式的观点——进化心理学和"角色理论"进行推论,假设男性和女性出于不同的原因背叛。玩 P/D 博弈的男性背叛,是因为他更具竞争性,更愿意去冒风险。这种动机称为**贪婪**。玩 P/D 博弈的女性会背叛,是因为她有更少的竞争性,更不愿意冒风险。这种动机称之为**恐惧**。P/D 博弈包含了恐惧和贪婪两者。因此,男性和女性会以相同的比率选择背叛。

辛普森创设了另外两种博弈——恐惧与贪婪,每一种仅包含相应的动机。两个博弈都没有占优策略。在贪婪博弈中,仅当对方合作的时候,背叛才会更好些。如果对方背叛,博弈者从合作和背叛中获得完全相同的收益。辛普森假设,在贪婪博弈中男性比女性更多地选择背叛。在恐惧博弈中,仅当对方选择背叛的时候,背叛才会更受偏好。如果对方合作,那么博弈者从合作和背叛中获得完全相同的收益。辛普森假设,在恐惧博弈中女性比男性更频繁地选择背叛。

现在我们可以看到实验设计中差异法的使用与理论的结合。博弈理论被用来创设了三种社会困境。性别差异是否贯穿于这些困境中呢?只有混合设计才能回答由下列原因产生的问题。虽然博弈理论创设了两难困境,却没有区分性别。事实上,博弈理论把所有的人看成以完全相同的方式决策。因此,对于发现性别差异是否存在假设的检验,差异设计是很必要的。

---

10 大量的策略已经被设计出来,用于重复的相互合作的 P/D 博弈,从而可能获得共同的更高收益。例如,在"以牙还牙"中,博弈者第一轮选择合作,随后采取前一轮对方所选择的选项,那么两个持"以牙还牙"博弈策略的人将一直合作下去。(Axelrod, 1984)

这些实验是在两种情境下实施的。情境 1:被试坐在个人电脑前,对着据称坐在另外一间实验室里的虚拟搭档进行博弈。被试被告知他们将与三个不同的搭档进行三个博弈,还被告知在实验里每得到 1 分就将获得 1 美元,但是他们不知道别人的选择或者总分,直到三个博弈全部完成。事实上,没有其他博弈者和被试因为参与实验而获得固定数额的报酬。第 2 个情境在两个方面有所不同:①被试在纸上给出他们的答案;②告诉他们想象这些分数对他们和其他人是有价值的,但是没有提及钱数。这两种情境在其他所有方面都是相似的,比如第 2 种情境下被试也将获得固定数额的报酬。

辛普森的研究使用了 3(博弈)×2(情境)×2(性别),12 个条件的**因素设计**。因素设计(Campbell and Stanley,1966)允许实验研究者同时对比多种差异。例如,**统计控制**可用于保持博弈和情境恒定,以至于在个人电脑环境下分析者能比较男性和女性对 P/D 博弈的反应。辛普森能够比较博弈(恐惧、贪婪及 P/D)、情境(个人电脑和非个人电脑)和性别(女性、男性)的效应。这些比较是因素的主效应(自变量)。因素设计也使得研究者可以评价交互效应问题,即两个或更多自变量的结合效应。事实上,辛普森的假设就涉及了交互效应(例如,女性和男性对不同博弈的反应是相同还是不同的)。

两种情境的结果部分支持辛普森的假设。正如假设所说,P/D 博弈的两种情境下男性和女性以相同的比率背叛。对恐惧博弈而言,任何一种情境下男女在背叛上没有差异。这个发现结果与辛普森的假设(女性更可能背叛)相反。最后,贪婪博弈中两种情境的结果都强力支持了辛普森的假设:男性比女性更可能背叛。

这一系列实验有很多好的设计优点,能轻易地满足经验驱动型实验的四条准则。第一,尽管辛普森没有随机分配被试到实验组和控制组,但是他使用了更多强有力的技术以确保情境的相似性:**所有**的被试完成所有的三个博弈,其中 P/D 博弈充当控制条件。因此,实验组和控制组的被试不仅仅是类似的(如同随机分配下所假定的那样),而且是相同的。第二,被试在不同的情境下进行同样的博弈。第三,在具体情境的博弈

中，除被试的性别外条件都是相同的。在女性和男性被试的反应上，博弈具体的差异不能归因于样本的异质性。因为异质性被排除了，所以我们能确保观察到的差异反映了博弈、情境以及性别上的差异。例如，贪婪博弈中在背叛比率上女性与男性的差异取决于博弈效应，而不是被试的特征。

辛普森运用博弈理论创设了 P/D、恐惧和贪婪三个博弈，考虑到每个博弈中可计算出的恐惧和贪婪动机的相对强度（例如 P/D 的差异是 0），然后精确地比较不同的博弈。三个同等重要的博弈定义了三种截然不同的社会两难困境，从而反映出发生在实验室外的情境。例如，我们已明确在社会内和社会间的合作问题上会频繁采用 P/D 博弈的形式（Axelrod，1984）。似乎毫无疑问，未来的研究将会发现采用恐惧和贪婪博弈形式合作的社会问题。

此外，在两种情境中得出的研究结论是强有力的，因而减少了特定情境产物效应的可能性。不像米尔格拉姆的实验，实验者在效应的产生上没有直接的作用；也不像阿施或者米尔格拉姆的实验，没有使用实验者的同伴。因为同伴无论得到多么好的训练，通过行为中无法注意到的变量，都会引入需求特征和实验者行为偏差，从而影响到被试的行为（见第 5 章）。既不是实验者，也不是同伴引入了对比的实验条件，博弈做到了这点。因为实验条件不受被试—实验者或者被试—同伴交互作用的影响，所以将人为产物被引入实验就几乎没有任何可能了。

评论综述

这里分析的四个实验表明差异法实验应该怎样和不该怎样建立。好的实验设计能发现自变量和因变量之间的联系，其方法是通过改变自变量出现与否，或者自变量的值，以观察因变量是否有相应的变化。实验条件和控制条件仅在一个方面存在差异。对研究中现象的看法，决定了哪个条件作为控制条件，哪个条件作为实验条件。在阿施的从众研究中，控制条件

排除了社会压力,而实验条件采用了不同水平的社会压力。辛普森运用 P/D 博弈作为其他两个博弈的控制条件,事实上,其他两个博弈中的任意一个仅在一个方面与 P/D 博弈有所不同。

这里评价的四个研究每个都是一系列补充研究的起点。阿施实验表明占多数人的数量和一致性影响从众,但这仅仅是一个开始。很多其他的因素也可能影响从众,其中包括相对于被试的同伴地位,大多数人反馈的可信性,等等。与没有地位差别的阿施被试相比,低地位被试是不是更可能遵从七个高地位被试的反应呢?如果是这样的,相对地位与多数人数量的大小结合起来是如何影响从众的呢?

米尔格拉姆实验在全球许多国家得到复制,并得出了相似的结果。他们最后证明,人类给自己同伴造成伤害的能力是真实存在的。此外,米尔格拉姆发现,与老师和学习者身体和心理上的接近等这样的因素也影响了服从比率。而且,最近美国政府官员和其他国家权威人士拷打普通公民的证据重新使人们强调这类研究的重要性。然而据我们所知,对服从原因有不同的解释——包括我们提出的这些——还没有被排除。它们没有被排除的原因是没有人实施米尔格拉姆的实验了,众所周知在这类实验中自变量是实验者的合法性,或者仅仅是实验者的合法性。一直声名狼藉的米尔格拉姆的设计与制度评审委员会日益严密的审查相伴相随,使得今天复制这些实验变得更加困难[11]。尽管如此,具有创新思维的人们将找到其他方式设计新的实验,这些实验将严格地设计控制组和实验组。到那时,合法的权威和强迫在服从上所产生的效应将得以区分,并且其他动机效应将被排除在外。

巴夫等人的实验设计得很好:他们满足了经验驱动型实验所有的 4 条准则。结果,他们采用了一个强有力的例子来说明:行为是自动形成的,无需有意识的感知或判断过程作为中介。自动化是否能产生研究中没有考虑到的一系列其他结果

---

11 所有研究型的大学和研究组织都有制度审查委员会(IRBs),负责评价和保护对被试的合乎伦理的处置。他们会在参与者的风险和研究结果的收益间作出权衡。第 5 章将讨论在执行伦理准则上 IRBs 所应具有的伦理和担当的角色。

呢？例如，即使在完全理性的情况下，决策也会受到决策前框架条件的影响。另外，与获利的正效用相比，同等数量的损失具有更大的负效用。自动化能框定理性的决策吗？如果是这样的，我们将认识到，前意识过程会不知不觉地框定人类活动的大部分意识，并框定决策的制订。

转向辛普森的研究，后续的研究工作已经完成了。库瓦巴瑞（Kuwabara，2005）提出，恐惧博弈的负面结果是由于辛普森运用了错误的博弈去验证假设而造成的。更好的验证应该设计一个博弈，实验中被研究的恐惧是对对方贪婪的恐惧！库瓦巴瑞创建并施行了这样的博弈：利益是对对方贪婪的恐惧的合成。实际上，他的实验被试是面对模拟的对方贪婪博弈，进行恐惧博弈。库瓦巴瑞发现如辛普森开始所预测的，女性背叛远高于男性背叛。

回顾这里评价过的研究，巴夫等人以及辛普森所做的两个实验设计得特别好，提出了令人兴奋的新发现。而阿施的尽管有些设计存在缺陷，但也得出了令人兴奋的新发现。阿施发现一致性程度与被试从众的程度是相关联的，还提出了产生从众的最佳小组大小的证据。

到目前为止，米尔格拉姆的实验是四个实验中设计最差的。他制造强力战术效应，这种效应随教师与受害人和实验者的接近程度而发生变化。但是服从是受到权威的影响吗？因为权威并没有从权力中分离出来（正如阿施分离了权力和影响力），它从未发生变化，从权威对服从的影响效应中我们什么也得不到。因此，米尔格拉姆所发现的合法权威与服从关系的主张是毫无依据可言的。

正如我们在本章开始时所提出的那样，设计良好的经验驱动型实验已经成为发现现象的强有力工具，这点可以从四个实验中的三个里得到很好的诠释。第一个到最后一个研究之间的五十多年无疑说明了实验设计的重大改进，这保证了未来更深入的发现。

这些发现的重大意义与实验发现能否应用于实验室以外有很强的关系。在应用经验驱动型实验的结果时，实验越贴近实际，越少有人为因素越好。贴近实际的实验更好的原因是，普遍

适用的发现要求实验条件与实验室以外世界的实例一一对应。一个新例子说明了这一点。卢卡斯,格瑞夫和洛瓦哥列(Lucas, Graif and Lovaglia,2006)假设,随着犯罪严重性的增加,控方的不当行为(包括隐瞒信息以防辩方知晓)也会增加。卢卡斯等人创设了控方极可能面临的条件,支持了他们的假设。假定实验与控告的条件相似,他们结论的普适性就变得很直接了[12]。

相比而言,我们很难想象像阿施所做的那些研究适用于实验外的条件。为什么阿施设计的实验带有很高的人为环境,从而阻碍了实验室外的推广呢?他这样做是为了①发现群体压力如何影响从众水平,即使被试明确地知道从众反应客观上是错的;②在如此稀缺的条件下,同伴的压力与随后从众之间的关系是明确的。一般来说,实验条件越简单,研究者对自变量和因变量之间的关系就越明确。然而,实验条件越简单,实验室外推广就越难。**对于经验驱动型实验,内部效度和外部效度恰好不能统一起来。**

与一一对应应用研究结论到实验室外的方法相比,米尔和他的追随者一直都需要一个更有效的方法。这种方法隐含着米尔的逻辑,但却没能发展起来。按照米尔的观点,实验是一种分析法,因为它们将世界分解成了更简单的部分。因此,米尔本应该设计一种综合方法,重新将规律结合到米尔假设的复杂整体中。然而,米尔和费希尔都没有设计出这种**综合方法**,之后也没有人提出这种方法。

下一个讨论的主题是理论驱动型实验——由理论所设计的研究。理论驱动型实验的逻辑创建了一种方法,即将世界分解为更小、更简单、更便于实验室研究的部分。它也包括了综合法,以解决重新整合实验结果的问题,以此来把实验和自然发生的现象联系起来。在物理实验室所发现的结果可以应用于数千光年外的星系研究,这种情况能否同样地适用于社会学呢?我们断言是这样的,下一章中表明了:理论驱动型实验方法在社会学与物理学上是完全相同的。

---

12 辛普森的实验是一一对应相似要求的例外。正如我们上面所说,他的结论更容易应用于实验室外,因为混合设计采用了先前被验证过和应用过的理论。

4 理论驱动型实验

在本章,我们会将实验设计逻辑的研究扩展到理论驱动型实验。理论驱动型实验的检验源于理论的假设,但理论的任务不仅限于提供假设。**理论设计理论驱动型实验**,此外,理论是连接受控实验室环境中进行的观察与实验室外的世界之间的桥梁。由于理论设计实验,因此在理论驱动型实验中没有一个设计(例如:控制组对实验组)可以包罗一切。相反,实验设计随着所检验的不同理论而呈现多样化。但是,所有理论驱动型实验都有一个共同的逻辑,相应的,也有共同的优秀设计的原则。

本章围绕着一个核心观点展开:**理论是科学的方法——所有的科学都是如此**。在第一部分中,我们将描述理论驱动型研究的逻辑基础,并明确优秀设计的原则。据我们所知,这些原则以前还不曾出现在书本上。我们通过分析两组实验开始:第一组是物理学实验,验证了阿基米德的杠杆原理,这是最早提出的科学原理;第二组实验验证简单的社会结构思想。通过这些实验,我们可以发现这两大领域实验所共有的优秀设计的原则。

如果没有进一步的证据支持,我们就此宣称所有的科学都使用同样的方法和同样的原则似乎空洞无力。我们再分析四个实验,展示一对分别来自社会学和物理学的结构相同的理论如何设计相同的实验。第一对是伽利略的自由落体理论和詹姆斯·穆尔的基于地位的影响研究。每个都是**发现式实验**(founding experiment),因为它们都:①根据理论设计;②验证该理论;③出现在该理论的科学系统发展的早期。后面两个实验

也是发现式研究,根据几何光学和网络交换理论进行设计。

理论科学和理论驱动型实验

理论科学以一种不同于 J. S. 米尔、R. A. 费希尔及其他追随其后的经验主义者的方式理解世界。理论科学不是围绕"世界是有规律的"这一无根据的假设而展开。因而,理论驱动型实验并不用于发现(假定的)潜在规律。理论科学也不是建基于"世界是无规律的"这一同样无根据的假设。那么理论科学在世界的规律方面持什么观点呢? 它声称,世界是否有规律不能仅仅根据理解世界的理论来判断。**经验表明理论可以使世界表现出规律**。理论是揭示了潜在的规律性,还是仅仅系统地表述了我们的理解? 没有人能回答这些问题。

正如第 3 章所指出的,尽管实验的方法不同,其目的却都是发现规律性,因此它们必须重复进行。之所以要重复实验,是因为单项观察的结果并不是规律。虽然理论驱动型实验的目的不是发现规律,但重复实验对于理论驱动型研究同样很重要。这些思考会产生两个问题:第一,在理论科学中,重复实验什么? 第二,由于理论驱动型实验的目的不是发现规律,而是检验理论,那么为什么理论科学需要重复实验?

科学家在验证同一个理论推导时会重复几次实验,通常会产生类似的结果。这些结果之所以类似,是因为理论本身已足够强大从而形成了规律吗? 还是重复实验的结果取决于理论中潜在的未发现的规律? 同样,也没有人能对此进行回答。它们也不需要回答,因为理论的重要性取决于它们的**确认状态**(confirmation status)以及理论所通过的一系列检验。反过来讲,它们的确认状态只取决于理论检验的结果,而与无法证实的关于世界本质的假设无关。

重复实验可以有第二重含义:在理论的应用范围内实施的一系列实验。这里的实验并不相似,但都会由理论推导所涵盖。这种类型的重复实验一旦成功,将对理论提供非常强大的支持。正如波普尔在他关于理论效力的讨论中所指出的:

可检验性有多种程度：一个理论越肯定地断言，冒的风险也就越大；与断言得少的理论相比，也更能被验证。[1]（Popper，1994：94）

重复实验之所以在理论科学中处于核心地位，有两个原因：第一，如果理论是正确的，实验会得出具有可重复性的研究发现。这样，重复实验就成了理论效力的重要证明。第二，重复实验具有核心地位，因为**科学知识是大众知识**。虽然提出一个理论或进行第一次检验的研究者，在发现支持他们理论的证据的创造性研究上，可能享有个人权益。但是，由于科学是大众的，一个研究者可以检验另一个研究者的工作。由于他人能够仔细检查旧的实验结果，并产生新的实验结果，研究将会更仔细、更如实地进行。作为一项规则，科学界会对得到重复实验支持的理论给予更大的信任，至少，其中的一些重复实验是由没有利益冲突、对于理论的"成功"无法分一杯羹的第三方完成[2]。

任何关于社会学和其他社会科学的重复实验的讨论，都有可能引起一定程度的怀疑，因为很少有这方面的重复实验。重复实验之所以这么少，部分原因是大多数的社会学研究是描述性的，或是建立在米尔的逻辑上的，这种方式难以进行重复实验。比如，调查提供的信息是来源于特定时间和地点的特定群体。调查结束后，所进行调查的时间已经过去，并不会再现。因此，调查不能被重复[3]，针对特定时间和地点的特定群体的案例研究也不能被重复。实验的批评家或许会指出从来没有一项实验能够以完全相同的方式进行两次，因此他们宣称理论驱

---

1 在波普尔的系统中，**可检验性**（testability）指的是一个理论被发现为错误的最大限度。一个理论越有机会被证据所驳倒，这个理论就越可检验、越强大（Popper，1994：88）。科学史上有许多例子，对于一些理论，起初人们无法想象或设计实证性的检验。但是，随着科学的进步，其中的许多理论得到了验证，并发现了支持（或证伪）他们的证据。

2 让我们来思考围绕冷聚变的"证据"展开的争论（Voss，1999）。结论无法复制可以反映出不好的理论、不好的研究设计、学术造假，或者三者皆是。在1989年，无论是报告冷聚变证据的研究者，还是其他人，都无法复制他们的结果。

3 检验理论的调查实验是例外，而且这些实验是由理论设计的。

动型实验不能重复。实际上,理论驱动型研究的重复实验并不需要研究者回到过去,而只需再次应用该理论。重复实验仅仅指理论以与过去同样的方式再次得到应用。

由于理论是科学的方法,理论驱动型实验必须与各门科学中的理论具有相同的逻辑关系。方法的统一会带来明显的优势:一个研究者可以通过研究其他科学中的理论驱动型实验,学会如何在社会学中设计和实施理论驱动型实验。也就是说,社会科学实验和其他科学实验有相同的教育意义。接下来,我们展示六个实验:三个物理学实验,三个社会学实验。理解这些实验不需要事先具备有关被验证的物理理论和社会理论的知识。在每一个案例中,我们会描述该理论,然后说明案例中是如何运用它设计实验的。

这六个实验将被成对介绍,其中一个来自物理学,另一个来自社会学,而且同一对实验比不同对的实验更有相似性。我们使用这些作比较是为了说明:①蕴含在理论驱动型实验中的逻辑;②不同科学所采取的实验形式都是相同的。理论驱动型实验在不同科学中具有相同的逻辑,这一事实具有许多重要意义。我们将随着章节的深入对这些意义进行介绍,但在这里先提两点。第一,在社会科学中经常能听到这样的批评声:实验室的实验结果不能普遍化。这样的批评是错误的,因为理论驱动型实验根本不是普遍化的。正如在前一章的结论中所解释的,这些实验的目的是为了验证理论。一旦得到验证,实际上是理论在实验室外产生影响,而并不是实验。事实上,社会学实验的结果与其他科学的结果以同样的方式应用于实验室之外(Lederman,1993:101)。我们怎么知道其他科学的理论可以应用到实验室外的地方呢?如果物理学和化学理论仅仅应用于实验室,那么你的车就无法启动,你的电脑将无法运行。事实上,你或者其他任何人都不会有车或电脑,因为它们就不会被发明或制造出来。

不同科学的理论和理论驱动型实验的统一,还有第二个意义:社会学中的科学进步过程与其他科学完全相同。科学通过依据更广泛、更精确的理论,提出更广泛、更精确的解释和预测,从而取得进步。科学进步最有效的方式是理论和实验间的

相互作用。正如本书其余部分的例子所说明的，社会学的一些领域已经在不断进步（同样见 Berger, Willer and Zelditch, 2005）。同许多其他科学一样，社会学中不断进步的研究领域通过理论和实验之间的相互作用得到提升。但是，我们必须再次说明，实验不是科学进步的必要条件，理论才是。在最近的遗传学变革之前，从达尔文开始，生物学就表明科学可以只通过很少的，甚至不通过实验而获得进步。生物学通过理论和比较法的相互作用获得进步，但是，可以证明的是，那样的进步比在理论驱动型实验中的进步缓慢而且困难。

实验设计的原理、模型和逻辑

现在是时候来揭示构建优秀的理论驱动型实验设计所必需的标准了。我们从阿基米德的工作开始。首先，我们设计实验来验证阿基米德的杠杆原理（Archimedes［公元前 232 年］, 1897）。以那些实验为例，我们确定设计优秀的理论驱动型实验的标准。然后我们设计实验来验证社会关系中的合理性原则，并将标准应用于它们。通过这些，我们可以提供第一个证明：用于构建理论驱动型实验的原理在物理学和社会学中是相同的。

验证阿基米德原理

古代伟大的数学家阿基米德（公元前 287—前 212 年）以富于创造性的发现和机械发明而被世人熟知。事实上，那些发现和发明来源于他几何学上的进步。例如，锡拉丘兹国王希尔罗要求阿基米德判断王冠是否是纯金的，看看金匠有没有欺骗他，在金子里掺了银。由于金的密度大于银，阿基米德知道解决方法就是测出王冠的密度。密度是物体质量与体积之比。王冠的质量通过天平很容易就能测量出来。但是王冠的形状是不规则的，怎样能不熔化它就知道它的体积呢？

在对这个问题冥思苦想了数天后，阿基米德在洗澡的时候注意到，当他进入浴缸时，有些水溢了出来。他突然想到，他的身体、王冠或其他任何固体在被水浸没时，会正好替换掉体积

相同的水。故事的结局是,阿基米德光着身子跑回家,并大声呼喊:"找到了!我找到了。"

关于杠杆原理,据说阿基米德对希尔罗断言,"如果给我一根足够长的杠杆和一个支点,我可以撬动地球"(Heath,1897)。事实上,阿基米德如此精确地得出杠杆原理,这一原理的应用在阻击包围锡拉丘兹的罗马人时起到了很大的作用。这些应用像理论的任何应用一样,验证了该理论。以下实验设计是对阿基米德的杠杆理论的验证。

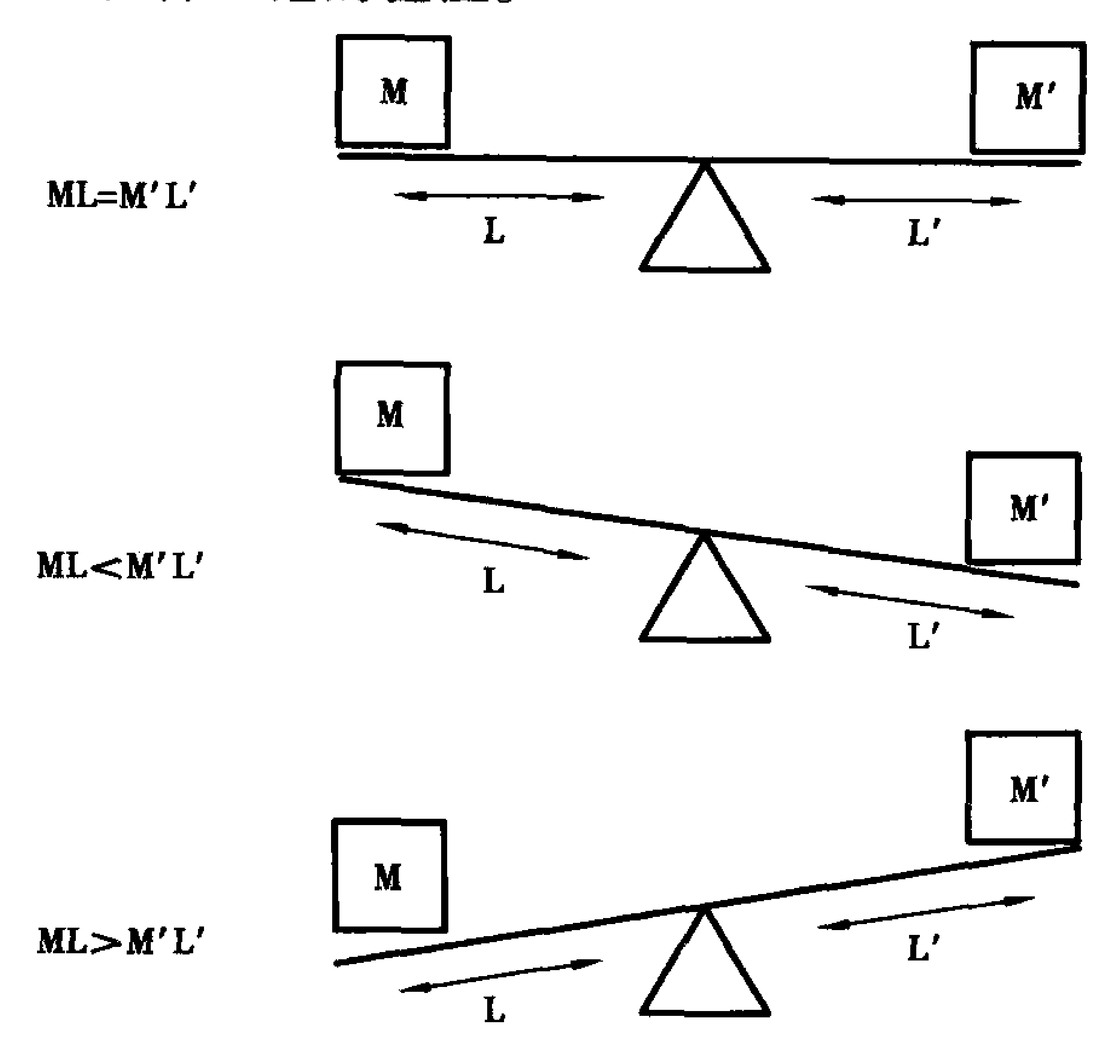


图 4.1　阿基米德的杠杆原理模型

图 4.1 中,左边显示的是杠杆原理,右边显示的是运用杠杆原理的模型。三个模型,每一个都有一根杠杆,轴心处于三角形支点的顶端。两个重物,M 和 M′,放在杠杆上,相对支点的距离为 L 和 L′[4]。三个模型中,L = L′:支点到两个重物的距离相等。三个模型代表系统的三种状态。在第一个模型中,两

4　此处 M 代表质量,当然质量与重量不是同一个概念。质量是一个物体中的材料的数量,而重量是在重力场中质量所施加的力。例如,一个给定的质量在月球上的重量,是其在地球上重量的三分之一。由于杠杆的两端重力相同,相互抵消,比较的只是质量。由此可以推知,一个在地球上处于平衡的杠杆,在月球上也会平衡。不过,当根据 M 和 M′的重量进行推论时,看起来会更清楚些。

个重物质量相等，由于距离也相等，得出 ML = M′L′。杠杆原理的经验表明：当 ML = M′L′，如模型所示，系统保持平衡。在第二个模型中，左边重物的质量轻于右边重物的质量。由于 M < M′，而距离相等，得出 ML < M′L′。现在经验告诉我们，系统的重心不平衡，偏右。第三个模型与第二个相反：M > M′，而距离仍然相等，得出 ML > M′L′。系统的重心不平衡，偏左。

理论设计实验，因此，一个实验就应当看起来并且被预期运作起来符合图 4.1 的模型。当实验装置建立得和理论模型一样时，弗里兹和塞尔（Freese and Sell，1980）称这个装置是模型的**复本**。如果复本的表现如同模型一样，就不需要人们像火箭科学家一样假设杠杆装置有两个理想的属性：无摩擦的支点和无质量的杠杆。如果杠杆是无质量的，就不会与 M 和 M′的质量相互作用。如果支点是无摩擦的，杠杆的状态无论是水平，还是倾斜向一边或是另一边，都会反映出 M 和 M′所受到的力。但可惜的是，世界上没有无质量的杠杆和无摩擦的支点。那么该怎么办呢？

答案在于运用杠杆原理去推断实验会受到摩擦力和杠杆质量怎样的影响。当摩擦力足够大时，即使质量改变，系统仍会保持其初始状态——平衡或不平衡。如果这样，观察会错误地断定原理被证伪。但是如果原理是正确的，当我们改变 M 或 M′时，会发现摩擦力的大小与质量有关。例如，将 M 的质量一点点地增加，我们会发现杠杆开始移动的阈值，从而找到摩擦力的大小。如果我们发现了摩擦力的大小，就可以通过不断尝试找到使它变小的方式。换句话说，由于不能创建无摩擦的支点，我们将设计与 M 和 M′的质量相比摩擦力很小的实验器械。

关于杠杆，杠杆原理认为，如果杠杆的一端重于另一端，验证 ML = M′L′的实验肯定会受到影响。系统会表现出类似 ML > M′L′或 ML < M′L′的情况。幸运的是，我们可以在实验之前测试系统。首先，通过测量确保 L = L′。然后移动 M 和 M′，检查平衡。如果杠杆平衡，横杆的质量均匀分布，理论断言实验不会受到影响。实验之所以不会受到影响，是因为对 M 和 M′增加了常量，不会影响原有的相等或不等。

随着模型被实验有效地复制，下面只保留初始条件，进行

三个实验。我们之前已经使 L = L′。第一个实验中,我们需要为 M 设值,使 M = M′,而第二个实验中使 M < M′,第三个实验使 M > M′。我们不使用天平来设值。因为天平是基于杠杆原理的,当杠杆原理得到证实后,我们才有理由相信和使用它。我们使用弹簧秤设定 M 的值。实验中,M 和 M′会被放在杠杆上,观察结果平衡与否。

正如上面所解释的,图 4.1 右面的模型给出了实验装置设计;被验证的原理是用 ML(M′L′)值之间的关系描述的表达式,位于图 4.1 左面;图中杠杆的三个状态是预测结果。当实验结果满足了模型,理论就得到了支持。对理论的进一步应用将不限于那些被验证的案例,不过,由于有那些案例,我们对预测就更有信心了。

我们第一次体会了理论驱动型实验的一个重要特点。实验结果并不会令人惊讶——至少实验者不会惊讶。他们也不应该惊讶——理论驱动型实验的目的并不是产生理论没有预期到的惊人发现,相反,理论在实验之前已经预料到了"发现"。实验的目的是为了知道理论的发现是否正确。如果我们身处一个完美的世界,就不需要任何实验去验证结果是否正确。但是,理论和应用理论的人容易犯错,因此理论驱动型实验**确实**会有所发现——只不过,实验的发现对于实验者而言,有时是快乐,有时却是失望。

理论驱动型实验的准则

理论驱动型实验的原理直接遵循上述实例的逻辑。我们将之概述为 5 条标准:

1. 验证源于理论的一个或多个模型。
2. 通过将初始条件和终止条件联系起来,运用理论进行预测。
3. 建立复本,设定初始条件,并观察终止条件。
4. 将预测与结果进行比较,以确定理论是否得到支持。
5. 可以非常自信地从理论中推断,实例在理论上非常接近于支持理论的实验——但是预测不仅仅局限于类似的情况。

这些标准背后所蕴含的一致观点是理论模型（或多个模型）与相关的实验复本（或多个复本）相互间应该尽可能类似。实验是对相似性的检验。当模型及其复本在初始条件下相似，并且在实验条件下也相似，以至于模型预测了复本时，这个验证就是成功的。理论也可以用来预测过程。在预测过程时，同样可以比较连接初始条件和终止条件的模型和模型复本的多个过程。

正如杠杆原理的实验一样，当实验次数达到两次或两次以上时，理论驱动型实验看上去会和其他实验有惊人的相似。这样，那些只接受过差异实验（difference experiment）教育的人会理解为，三个杠杆实验在追求统计差异（empirical difference）。事实远非如此。实际上，基于以下理由每个杠杆原理的实验都可以单独成立。

一个简单的实验可以满足所有这 5 条标准。例如，设定 ML = M′L′，并观察杠杆是否像预测的那样保持平衡，这就是理论的验证。这之所以是验证，是因为**理论驱动型实验仅仅关注验证理论模型和实验复本之间的关联性——在理论的预测和实验结果之间**。理论驱动型实验不关心统计差异。那么为什么许多理论的应用需要进行多次实验，而且各个实验之间只是某一方面存在差异，就像三个杠杆实验一样呢？这是为了增加验证实例的数量，从而减少这种可能性——实验结果是由随机事件产生的吗？在理论的适用范围、概念的精确测量以及以简单模型表达复杂现象的精要程度中，可以找到这些问题的答案。

多次实验的主要目的是考察理论可以有效应用的范围。看一下杠杆原理：现在假设理论的适用范围仅限于实例 ML = M′L′。一个理论假设只有当 ML = M′L′时才有效，其效用是有限的。而一个理论在图 4.1 的三种情况中都有效，则它有更大的效用。当然，杠杆理论的适用范围不仅限于所讨论的三种情况。但是，如果我们已经将理论成功地应用于三种情况，而不只是一种情况，那么在它被用于更多的实验时，我们就会更自信，即使是用于不同于图 4.1 的其他情况。

下面描述的穆尔实验更进一步阐述了为什么一些理论驱

动型实验的设计类似于差异实验。在一些情况下,对理论的概念的测量是相对模糊的。以穆尔已经检验过的模型为基础的实验,预测了地位和影响之间的次序—层级关系。没有具体的预测,理论的一些意义需要多次实验来验证。因此,穆尔进行了多次实验,这使他能够在实验间进行比较,并确定理论的预测得到了支持。最后,由于一些设计有效地简化了复杂的现象,不得不进行多次实验,再次体验复杂性。每个穆尔实验都仅仅关注不同地位交互的一方(地位高或地位低的对象)。多次实验,有些针对地位高的参与者,有些针对地位低的参与者。实验后需要将实验发现与对不同地位的参与者之间交互情况的预测进行比较。

重复实验也很重要,因为单个实验的结果可能是偶然的。所谓偶然,意思是有些不被理论所了解的条件引起了预期的结果或者使预期结果无法实现。幸运的是,结果的偶然性会随着验证的增加而迅速减少。因此,随着实验次数的增加,我们对于理论的信心也大增。通过多次实验进行检验的最好方法是,每次只改变一个初始条件,这样不仅增加了验证结果的次数,还可以对理论的适用范围进行系统地考察。

在总结关于理论驱动型实验原理的讨论时,我们来看一下第五条标准:它认为,当在理论层面上预测的情况与在实验中考察的实例相似时,我们是最有信心的。但是,预测的范围不仅限于那种相似性。那些对理论不太了解的人会惊奇地发现理论的应用不仅限于已被验证的一系列现象。进一步而言,如果应用的范围如此有限,理论就会失去许多价值。

理论为拓展我们的知识提供了可能性,从"我所知道的"扩展到了"我想知道的"。在我们为什么建立理论的多个原因中,这或许是最重要的一个。不管怎么说,在某种程度上,将理论应用于它被验证的范围之外或它本身的范围之外似乎是一种危险的冒险。但是,还有什么选择呢?其他的选择就是没有理论地盲目行进。比较好的选择是依据已被验证的理论。经历了多次验证的理论可以有效地迁移到新情境中,即使新情境在它的理论范围之外或者它此前被成功验证的范围之外。当理论的应用范围被拓展了,除了解释或预测之外,每一次成功

的新应用都反映了理论已得到新的验证。这,正是科学知识的积累。

将标准应用到社会学实验

从古代哲人(Aristotle[公元前347年],1962)、中世纪的学者到近代经典社会理论学家如马克思、西默尔和韦伯,他们都曾面临这样一个大问题:社会结构是如何影响人类活动的。我们使用网络交换理论(Willer,1999)来探索混合动机的交易场所中交换结构和行为之间的关系。网络交换理论(Network Exchange Theory,NET),运用理性原理——所有的社会参与者都寻求最大化他们所期待的、偏爱的状态选择(Willer,1999:30)——形成图4.2的模型。这个原理表明①人们已预设了偏好;②人们的行为目的在于使那些预设的状态被排在最高等级;③因此,通过我们设计的模型可以推断他们的行为。我们首先介绍连接着A和B的混合动机社会关系(mixed-motive social relation)。然后我们将展示在那个社会关系中,A的行为和B的行为如何受到结构差异的影响,这种差异存在于提供给A或B专用的选择中。

混合动机关系是指两个参与者受到两种相反动机的驱动。当A和B可以达成很多协议时:①任何协议如果对A更有利,则对B更不利,反之亦然;②对于A和B而言,协议比冲突(即他们无法达成协议时)好。因此,参与者之间发生竞争,是因为协议对一方有利,对另一方不利。参与者之所以合作,是因为对于双方而言,协议比冲突好。图4.2显示的三个模型,每一个都由混合动机的A—B关系构成,有一个价值为10的资源库,以及另外两个处于关系外部的、专属的收益。"专属"于A的意思是,A可以与B达成协议或者另外选择接受收益(显示为A左边的箭头),但不能同时兼得,B亦如此。有了这些思路,我们可以使用模型来建立实验复本了。

在所有的三个模型中,如果A和B同意分享,他们就分到价值为10的资源。但是如果他们不能对资源的分享达成协议,他们就不能在此关系中获得资源。理性原理表明:A和B想要更多而不是更少的资源,他们谁也不会接受得不到任何资

源的分享方式。下一步,我们规定资源可以以 1 为单位进行分享。也就是说,在第一个模型中,A 和 B 都没有其他的收益:那么就有 9 种可能的协议,对于 A 而言就是 9 ~ 1,而对于 B 而言,就是 1 ~ 9。任何协议,如果 A 获得的多,B 就获得的少,反之亦然。在这个范围内的所有协议都好于没有协议,因为那样的话,A 和 B 都不能获利。因此,第一个资源库关系是混合动机。其他两种关系也是混合动机,但是如我们所见到的,协议的范围由于其他的收益而变小。

$P_A=P_B$	0 → A —10— B ← 0
$P_A>P_B$	5 → A —10— B ← 0
$P_A<P_B$	0 → A —10— B ← 5

图 4.2　交换结构的对称模型和非对称模型

图 4.2 右侧模型的预测获利如下:第一个模型结构是对称的,而且,两个参与者是对等的,因为他们仅仅由理性原理所定义。他们的获利(P_X)必须相同,因此 $P_A = P_B$——A 和 B 各得到 5。在第二个模型中,A 的专属收益是 5,而 B 是 0。根据理性原理,A 不会接受任何与 B 达成的少于 5 的协议,仅仅提议 5 对 B 而言是一个冒险,因为对 A 来说, B 提议的 5 与来自于关系之外的收益没有区别,因此,A 可能会选择后者,使 B 不能获得任何收益。更确切地说,当两个选择相同时,一个纯自利的 A 会随机选择,这样使 B 只有一半的机会。为了确保不会零获利,B 必须至少提议 6,由于有 10 可以分享,B 的获利不会大于 4,因此,$P_A > P_B$。第三个模型只是与第二个模型相反。由于 B 现在有非零的选择,为了避免零收益,A 必须提议不少于 6,获得不多于 4,因此,$P_A < P_B$。

理论设计实验。因此,实验会设计得看起来像图 4.2 中的模型,并且预测实验会表现得像模型一样。实验建立的应该是模型的**复本**。如果我们的复本行为表现如同被建模了一样,那么我们就不需要马克斯 · 韦伯来推断需要两个理想的质点。实验系统应该是:①完全独立,没有直接来自于模型系统外部的收益;②被试(被我们设定为 A 和 B 的)没有“额外”收益。“额外”收益是一个诱惑,被试用于使另一方接受自己的提议。

外部获利和任一方的"额外"收益的获利都必须受到控制，因为它们不是模型的一部分。如果没有外部获利，且任一方没有"额外"收益，那么在模型中，A 和 B 都受到限制只能分享 10 个资源或接受替代的收益安排。世界上不存在完全独立的系统，也不存在完全理性的参与者在有选择权利的情况下放弃"额外"收益，以使他们之间的协议更好。那么我们该做什么呢？

答案就在于使用理性原理去推断实验会如何受到没有预先纳入模型的收益的影响。理性原理表明只有当已知所有相关收益时，活动才能被预测。例如，如果模型之外的盈利与模型中的相比足够大，那么它就会被接受，A 和 B 就不会达成协议，他们谁都不会接受模型中的那一个收益。但是理性原理表明，在三个模型中，A 和 B 会交换。如果他们不交换，没有达成协议就会成为质疑理性原理的理由。因此，理性原理的应用表明模型的验证需要研究者尽可能创建一个完全独立于外部收益的模型。

怎样建立最好的复本呢？为了消除 A 或 B 可能引入的"额外"盈利，调查者要严格监控他们的交流，确保被试不能面对面。但是正如在第 3 章所发现的，让实验者直接与被试交互并不可取。因此，让实验者监控交互的更好方法是，创建一个环境，让所有的交互都通过计算机界面进行。计算机界面限制了被试的交流，使他们只能提供资源或达成协议。

原则上，社会科学实验没有理由不像物理科学实验一样精确，甚至比物理科学实验更精确。在杠杆原理的验证中，不可能创建一个无摩擦的支点。然而，上文刚刚描述的实验设计则可以被称为"无摩擦"，因为其复本准确地对应模型。可以证明的是，这些设计与用于验证杠杆原理的设计相比，能够更精确地复制模型。

建立完计算机界面，还需要激发被试始终遵循理性原理进行实验。为了激发他们偏好更高或更低的收益，实验者根据被试在实验中的收益来付给他们报酬。加上这样的条件，实验要求他们寻求最好的收益。在实验中，被试被安置在分开的房间中，每一个被试都会通过协商确定选择 A 角色或 B 角色（为了避免可能的偏袒，计算机系统会让所有被试在他们的伙伴处在

B位置时,都能看到自己处在A位置)。每一个设计都有10次尝试,以及与模型作出的预测相对应的实验结果。在第一个实验中,预测是$P_A = P_B$,实验结果是A和B平分各得到5点;在第二个实验中,预测是$P_A > P_B$,实验结果是A的收益比B多;在第三个实验中,同样如预测,实验结果是B的收益多于A。

在第二次实验中,我们发现实验结果并不出人意外——它们也不应该出人意外。在实验开始前,预测就已经得到确信了。因此,除非是理论有错误或者是实验设计有错误,否则实验结果应该是预测的结果。实验没有产生新发现,研究也没有期望实验有新发现。关于这一点,在所有的科学中,发现是重要的,但是在所有理论驱动型实验中,发现就是模型的预测。

实验的目的是为了验证理性原则能否成功应用于图4.2所呈现的模型。现在,我们来回顾一下实验设计,并评价一下他们是否符合优秀设计的5条标准。下文中,标准用**加粗字体**表示:

1. **验证源于理论的一个或多个模型**。建立了三个模型,每一个都与图4.2所示的一个模型相对应。

2. **通过将初始条件和终止条件联系起来,运用理论进行预测**。模型给出了初始条件,并运用合理性原理产生了三个预测。它们显示在图的左侧。

3. **建立复本,设定初始条件,并观察终止条件**。以计算机为媒介进行建构,会使复本非常接近于它们的模型。实验开始进行时,要先设定与模型相关的条件;在实验期间,这些条件保持不变;在实验结论中,观察A和B的收益。

4. **将预测与结果进行比较,以确定理论是否得到支持**。每次实验的结果都与预测非常相符。理性原理在模型中的应用得到了支持。

5. **可以非常自信地从理论中推断,实例在理论上非常接近于支持理论的实验——但是预测不仅仅限于类似的情况**。

一致的想法是:每个模型及其复本如果在最初时相似,终止条件也会相似。理论验证的就是这种相似性。这里验证支持了理论。理论之所以得到支持是因为每个复本及其模型初

始条件类似,其结果也类似。那些没有看到经验驱动型实验和理论驱动型实验之间区别的人,或许会去检验三个复本之间的结果是否各不相同。虽然这样的验证并无害处,但作为理论的验证,它的结果与实验无关。

就第五条标准而言,理性原理可以应用到许多模型中,它们不同于我们已研究过的三个模型。例如,有一个结构类似于图 4.2 中的第二种情况,但其中 A 与 C 有第二个资源为 10 的关系,这取代了固定选项。现在 A 有两个专有选项,可以与 B 或 C 交换,但不能与两者同时交换。理性原理预测 B 和新来的 C 之间会有一场出价的战争。为了达成协议,C 必须出价高于 B,同样,B 也必须不低于 C。无论第一次出价是什么,出价过程的最终结果是 9—1 会提供给 A。因此,理性原理能预测 9—1 的资源分配,但不能预测 B 或 C 哪一方会赢取与 A 交换的机会。这个应用是对第二种实验的简单扩展。因此,我们应该对预测有信心。

还有一个改动的步骤是这样一个模型:允许 A 给 B 发送一个重大的负面制裁。通过多次尝试,如果 B 还是不同意提供 A 所想要的,一个理性的 A 会发送制裁。在图 4.2 的第一个模型中添加那个负面制裁,模型就不再对称。A 有优势,理性原理预测 $P_A > P_B$。这个新结构是在已经完成的研究中改动了两步得到的,与预测 A 有两个专有资源库关系的资源分享相比,我们没有信心预测添加了负面制裁后的影响。不过,运用理论作出新的预测还是比没有理论根据的"凭经验猜测"好。

自由落体和基于地位的影响

前一节定义了理论驱动型实验的 5 个原理,我们把它们表述为标准。从验证杠杆原理的实验可以推论,这些原理也可应用于社会学实验——运用了三个简单结构模型来验证理性原理。在本节中,我们再提供两个实验例子,并根据 5 条标准来评价它们。第一个实验来自物理学,伽利略的自由落体研究;第二个是穆尔的地位效果研究。由于这两个实验都是理论驱

动型实验,如果它们设计得好,应该相互类似,至少都可以达到满足5条标准的程度,甚至还会有更多类似情况。

自由落体的研究

自由落体问题是指物体在重力的作用下自由落下的问题,这是一个从古代就困扰科学家的问题。亚里士多德的理论只关注下落的速度,他认为下落的速度随物体的**自然下落速率**(natural rate of fall)的变化而不同。在他的理论中,自然下落速率是物体的固有属性,即与物体的重量成正比,与下落时物体所处的媒介密度成反比(Aristotle[公元前330年],1961:162)。他的理论在伽利略之前被广泛接受。

伽利略(Galileo[1636],1954)用两种方式击溃了亚里士多德的理论。第一,他运用亚里士多德的理论设计了一个实验:伽利略从300英尺的高度扔下一个200磅的炮弹和一个小于0.5磅的步枪弹。应用亚里士多德的理论,炮弹应该比步枪弹的下落速度快400倍(400=200÷0.5),但事实并非如此。伽利略观察到炮弹比步枪弹先到达地面,但前后差距不超过一英尺(Galileo[1636],1954:62)[5]。他证明亚里士多德的理论有经验错误,但这个证伪的工作还只是开始。

接下来,伽利略开始第二阶段研究,他要验证亚里士多德的理论不符合逻辑。为了证明这一点,他提出下面的**思想实验**(thought experirment)。首先,从亚里士多德的观点出发:重的物体自然会比轻的物体下落得快。接着,做一个混合物体,让轻的物质在上:在下落时,轻的物质产生向上的力,应该会阻碍重的物质下落;重的物质产生向下的力,应该会加速轻的物质下落。因此依据亚里士多德的理论预测混合物体会以介于两个物质单独下落速度之间的速度下落。但是,由于混合物体比它的任何一个组成部分都重,一定下落得更快,因此这个预测不正确。因此,亚里士多德的预测自相矛盾。至此,伽利略证明亚里士多德的理论经验上有错误,逻辑上也不相符。推翻了亚里士多德的理论,伽利略就需要提出更好的理论。

---

5 据说,伽利略当时借助了比萨斜塔进行实验。

提出新理论的第一步是将注意力从持续关注了2 000年的自然速度转向**自然加速度**(natural acceleration)。自然加速度是指下落物体的速度**变化**率。伽利略的理论是所有自由落体的加速度是一样的。为了作预测,他提出**匀加速运动**(uniformly accelerated motion)的概念。“所谓匀加速运动,是指速度从零开始,在相同的时间间隔内,速度改变量相同”(Galileo[1636],1954:162)。举一个匀加速运动的例子:设 x 为距离,物体在1秒末的下落速度为每秒 x,2秒末的下落速度为每秒 $2x$,3秒末的下落速度为每秒 $3x$,以此类推。例如,今天我们知道地球的重力加速度(原文此处为“重力”,有误)约为32英尺/秒^2,根据伽利略的理论,第1秒末物体的下落速度为32英尺/秒,第2秒末为64英尺/秒,第3秒末为96英尺/秒,以此类推。

回想前面的例子,实验者经常设计新的测量方法来研究现象。伽利略在尝试开展实验验证他的理论时,面临一系列测量问题。最简单的验证是要测量物体在某一瞬间的下落速度。在3秒后自由落体是以每秒 $3x$ 的速度下落吗?为了回答这个问题,他应该先测出瞬时速度,即在一瞬间的速度。这样的测量直到20世纪发明雷达(后来是激光)之后才有可能进行。另一方面,和我们现在的做法非常接近,通过沿着一条直线设置标准的 x 倍,就可以测量距离。因此,首先需要拓展匀加速运动的思想,将时间与下落的距离联系起来,才能够用实验验证理论。

伽利略的推理如下:由于加速度是恒定的,在 t_0 到 t_1 的时间间隔内,平均速度是开始时0和结束时每秒 x 的平均值。同样对于时间间隔 t_1 到 t_2,平均速度应该是每秒 x 和每秒 $2x$ 的平均值。同理可知,设 V_i 是一个时间间隔内的平均速度,那么 t_0 到 t_1 时间间隔内的平均速度以及 t_1 到 t_2 时间间隔内的平均速度为:

$$V_1 = \frac{\frac{0}{1} + \frac{x}{1}}{2} = \frac{x}{2}; V_2 = \frac{\frac{x}{1} + \frac{2x}{1}}{2} = \frac{3x}{2} \qquad (1)$$

单位是每秒某距离。第一秒末的距离是 $x/2$,第二秒末的距离是 $x/2 + 3x/2 = 2x$,其余类似。更普遍的是,用t代表时

间,伽利略认为 $x \propto t^2$(下落的距离与时间的平方成正比)。

对伽利略来说,剩下的事就是要测量时间,但测量时间也是一个麻烦的问题。所有的机械钟,包括惠更斯做的钟,都基于伽利略的钟摆运动理论。而惠更斯直到伽利略著书一个多世纪后才发明了第一台机械钟。没有机械钟,伽利略设计了一个替代方案:

> 为了测量时间,我们在高处放置了一个盛水的大容器;在容器的底部焊接一根细管,可以射出很细的水流,我们就收集流出的水,然后称重。(Galileo[1636],1954:179)

伽利略的解决方法依据的是:每个时间间隔与收集的水的重量成正比。一个水钟理论上是精确的,但是对于极短的时间却不精确。怎么可以利用水钟为快速下落的物体精确计时呢?由于没有设计出更好的钟,就只得使这一现象放慢速度。自由落体运动得太快,但物体在斜面上下滑的加速度却可以通过调整平面的垂直度变快或变慢。最后,伽利略设计了木制模型,模型中间有一个打磨过的槽,一个圆形的铜球可以在里面滚动。伽利略用下面一段话报告了自己的研究结果:

> 接下来,我们尝试其他距离,将通过全部长度所用时间与通过一半长度的时间、通过三分之二长度的时间,通过四分之三长度的时间、通过每一段长度的时间相对比。在这样的实验中,重复了 100 多次,我们发现,物体滚落的距离间隔是时间的平方,这对于所有倾斜面(有角度的面)都适用。(Galileo[1636],1954:179)

正如我们今天所说的,一系列的实验对伽利略的落体匀加速理论提供了强有力的支持。

现在我们运用 5 条标准来考察伽利略的实验。各个标准以**加粗字体**呈现:

> 1. **验证源于理论的一个或多个模型**。在匀加速运动思想的基础上,伽利略设计了一个模型测量落体的下落速率、距离和时间。
>
> 2. **通过将初始条件和终止条件联系起来,运用理论进行预测**。今天我们可以做很多实验得出 g,即重力加速度,把它用于方程 $x = 1/2gt^2$,来预测其余实验的结果。伽利略

没有得出 g，而是运用理论预测基本模型的依次重复的结果。由于他的理论认为距离随着时间的平方而变化，因此他预测落下的小球在相同时间间隔内经过的距离是不同的。

3. 建立复本，设定初始条件，并观察终止条件。 伽利略的复本被他测量时间的方法所局限。为了减慢过程，伽利略的设计呈现的不是自由落体，而是在斜面上下滑的物体。根据他的理论，平面尽可能平坦，加速度就会保持一致。初始条件包括近乎无摩擦的平面、平面的长度和斜面的角度。

4. 将预测与结果进行比较，以确定理论是否得到支持。 结果印证了预测：与伽利略的原理相符，在每个设定的初始条件下，每个被观察的距离都与时间的平方成正比。

5. 可以非常自信地从理论中推断，实例在理论上非常接近于支持理论的实验——但是预测不仅仅限于类似的情况。

考虑第五条标准，一系列重要的研究扩展了这些实验。伽利略作了很充分的准备来从他的斜面研究结果推导自由落体运动。通过显示物体下落的时间随着平面的倾斜而增加，推论得到了支持。当然，假如他的研究没有因为无法测量短时增量而受到限制，他可能会发现平面加速度接近于垂直面上的自由落体。在所有案例中，研究的下一步是预测抛射物的路径。为此，伽利略将抛物运动看作两个部分的组合，垂直方向是匀加速的运动（因为抛射物是自由落体），水平方向是速度恒定的运动。将两个运动结合在一起形成了抛物线——或一系列抛射物的抛物线。下一步，后来由牛顿完成，并将它拓展到开普勒的行星运动理论，但那是另一个故事了（见 Willer，1987：12）[6]。

---

6 伽利略的实验设施一直保存到今天。人们可以在意大利佛罗伦萨的科学博物馆看到它。那些对科学史感兴趣的人，可以从他的工作发展中看到现代科学思想的核心（见 Galileo，[1636]1954）。对于简单的自由落体，他没有画任何图示。但是随着伽利略的兴趣转向抛物运动，以及其他更复杂的现象，他画了一系列图示。早期的图示是实物式的。例如对木杆受力的分析，他画了一个木制的横杆嵌在石墙内，举起一个很大的重物。到了后期，图示的写实性逐渐淡化：它们变成了简单的几何模型。

基于地位影响的实验

社会影响过程从古至今一直受到关注。还记得前文提到过,亚里士多德认为落体具有自然下落速率,这是其内在属性。他还将影响力归因于受影响的人具有或多或少的易感性,这同样也是其内在属性(Aristotle[公元前 347 年],1962:199ff)。他指出,易感性是每个人的常态或自然属性。如果是这样的话,一个人对于影响的易感性在不同社会情境中将是一个常量。

现代社会学家质疑亚里士多德的观点。从阿施的实验中,我们可以看到随着同伴数量的变化,从众性在发生变化。从技术上讲,阿施的实验没有证伪亚里士多德的观点,因为正如第 3 章所述,从众性决定于权利而不是影响。现在我们要分析的实验正是要证伪亚里士多德的观点。这个实验依据的理论产生于第 2 章所讨论的贝尔斯实验,它是关于影响的。贝尔斯(Bales,1950,1999)的研究发现一个人能够改变他人信念,进而改变他人的行为的程度随着他们自身在社会结构中的相对地位而变化。与亚里士多德所不同的是,A 受到 B 的影响程度在不同社会环境中不是恒定的,而是随着 B 的地位而变化。更普遍的是,马克思、韦伯和西默尔将现代社会关系概念设想为一个情境,每个人都根据情境中其他人的行为,定位自己的行为(Weber[1918],1968:26)。他们关于社会关系如何影响行为变化的观点深入于当代关于权力和影响的观念中。

现在我们再次回顾发生在社会关系中的影响与权力之间的差异。例如,行为发生变化是因为信念改变了。若 B 改变了他的信念是由于 A 的作用,而由于那些改变了的信念,B 的行为发生了变化,那么 B 的行为就是受到了 A 的影响。权力通过认可影响了行为,但信念没有变化。当 A 通过一系列积极或消极认可改变了 B 的行为时,便产生了权力。通过施加积极或消极的认可而行使权力,但认可不是影响过程的一部分。

然而,最近的实验研究表明影响也可以产生权力事件(Thye,Willer and Markovsky,2006)。因为毋庸置疑,影响和权力在很多方面都相关,现在只是了解到了一些。任何关于影响的研究都必须控制与权力的关系,并尽可能消除它们的关系。

许多社会影响的研究都关注任务组的行为。任务组是将多人组织起来完成一个或多个任务的团体(大多数委员会都是任务组)。拓展研究发现那些特定社会类别(例如,男性或白种人)的任务组成员影响力大于其他类别的成员(例如,女性或黑人)。例如,混合性别的陪审团总是会选择一个有职业的男性白人做陪审团团长(Strodtbeck,James and Hawkins,1957)。在发明地位特征理论之前,这样的模式很难被解释(Berger, Cohen and Zelditch,1966,1972;Berger, Fisek, Norman and Zelditch,1977)。我们现在确定基本的SCT(从现在起我们使用SCT来表达地位特征理论)概念,显示理论如何用于设计实验和产生假设。

约瑟夫·伯杰和他的同伴开发了SCT。他们认为像性别、职业类别的属性是**地位特征**(status characteristics)的一个例子,这个概念由休斯(Hughes,1945)引入。下一步,他们通过识别和定义**特定**和**分散**的地位特征来提炼观点。特定的地位特征,像音乐能力,是属性或个人特点。不同状态(例如,高能力和低能力)有不同的社会价值与地位,而每个状态都与其特定行为的期望状态相关,并与状态排名相一致。例如,有高音乐能力的人被**期望**比低音乐能力的人歌唱得更好,乐器演奏得更好。

分散地位特征是从种族或性别角度来区别的,指:①状态(如男性或女性)排名;②每一个状态与一个或多个特定特征状态的关联(如女性被期望比男性更会抚育,但算术不如男性);③每个分散特征状态与一个**普遍表现期望**(general performance expectation state)相关联,并与状态问题相一致。举一个普遍表现期望状态的例子:在当今美国社会,女性一般被认为不如男性能力强[7]。

SCT解释为什么任务组(例如,陪审团或工作组)中成员的影响排名会与他们所拥有的地位特征的状态的社会排名相关。

---

7 在这里,有几点必须加以说明:首先,SCT强调这些标准用于识别分散特征;第二,这些是关于特征的**观念**,不应该被看作个人或团体的固有属性;第三,无论哪一属性(如性别),实际上是否是分散特征,都具有文化上的独特性,因此时间和地点也是这样。在当代美国社会,女性的地位往往比男性低,但是在瑞典,男性和女性是平等或接近平等的。如果瑞典(或其他国家)的女性和男性具有平等地位,那么在那里性别就不是分散特征。

理论的应用范围受到一系列明确定义的状态范围(SC)控制。

SC_1：两个参与者(P 和 O)需要完成一项有价值的协作任务——T。

SC_2：P 和 O 相信成功完成 T 取决于具备与任务相关的特定能力。

SC_3：P 或 O 不知道是否具备了成功完成特定任务的能力。

SC_4：P 和 O 处于具有分散状态特征的不同状态。

第一种状态有两部分：①任务结果必须在一定程度上是有意义的，这样小组成员就会考虑成功或失败的可能性；②任务是协作性的，需要小组成员共同完成(合作)。设计第二个状态的范围可以消除在情境中小组成员认为成功完成任务不需要技能或技能与之无关(如成功是一种侥幸)的情况。第三种状态需要小组成员不了解他们的技能。如果他们知道他们有相关任务技能，他们就不需要在任务情境中使用其他信息(如地位信息)。最后，第四种状态范围需建立一个最简单的地位情境[8]。

SCT 提供了 4 个理论状态(TS)或论据，解释为什么地位特征的排名与影响的排名相关，即使在场景中一开始特征似乎与任务无关。

TS_1：如果小组成员有不同的地位特征状态或如果他们具备的特征与任务相关，他们会激活(关注并使用信息)他们所具备的特征状态。

TS_2：如果成员考虑到他们的地位排名，他们会认为特征状态与决定任务成功或失败的能力状态相关。

TS_3：如果成员相信状态排名与任务能力相关，他们会期望能力状态以他们的地位排名准确分布。即地位高的被期望具备高的任务能力状态。

TS_4：期望具备高的任务能力状态的成员会比期望具

8　理论家们采用共同的策略。理论工作首先从最简单的可能情境开始。最简单的情境很少存在于现象世界。两个人在单一地位特征上存在差异的情境，几乎与无摩擦杠杆支点或无重力横杆一样罕见。

备低的任务能力状态的成员更有影响力。

SCT 的测试必须考虑到权力和影响力在社会关系中有同样的效果。正如伽利略多次使用光滑的倾斜面来减少他的落体研究中的摩擦力，研究者研究影响也必须排除实验情境中的权力。可以通过消除任务中的相反利益或消除被试相互认可的可能性来排除权力。

詹姆斯·C. 穆尔(Moore，1968)运用 SCT 设计了一个实验，实验中男性被试相信他们与另一个女性被试一起合作。他给了他们一些信息：他们和虚拟的合作者处于分散状态特征的不同状态(教育地位)。下一步，穆尔创建了一个任务，要求被试判断矩形盒子中被阴影覆盖的正方形是否多于没有被阴影覆盖的正方形。实验者告诉被试，他们必须尽可能作出正确的判断，并使用任何可用信息来帮助他们作出正确判断。

设计这个判断任务是为了研究信念改变时的影响，不是信念不变时基于权力的从众。认真看图 4.3，似乎很难作出准确判断。事实上，判断任务故意设计得模糊不清，这样被试看到的黑色(阴影覆盖的)正方形介于 45% ~55%。任务被设计得很难，但不是不可能完成。因此，当被试的反应表现出从一个判断转变为另一个判断，那样的改变会反映出信念的改变。相比较，阿施实验中线条长度的差异是明显的，被试反应不是由于信念的改变。因此，阿施研究的从众是由于权力，但穆尔研究的是影响。我们现在看看影响是如何施加的。

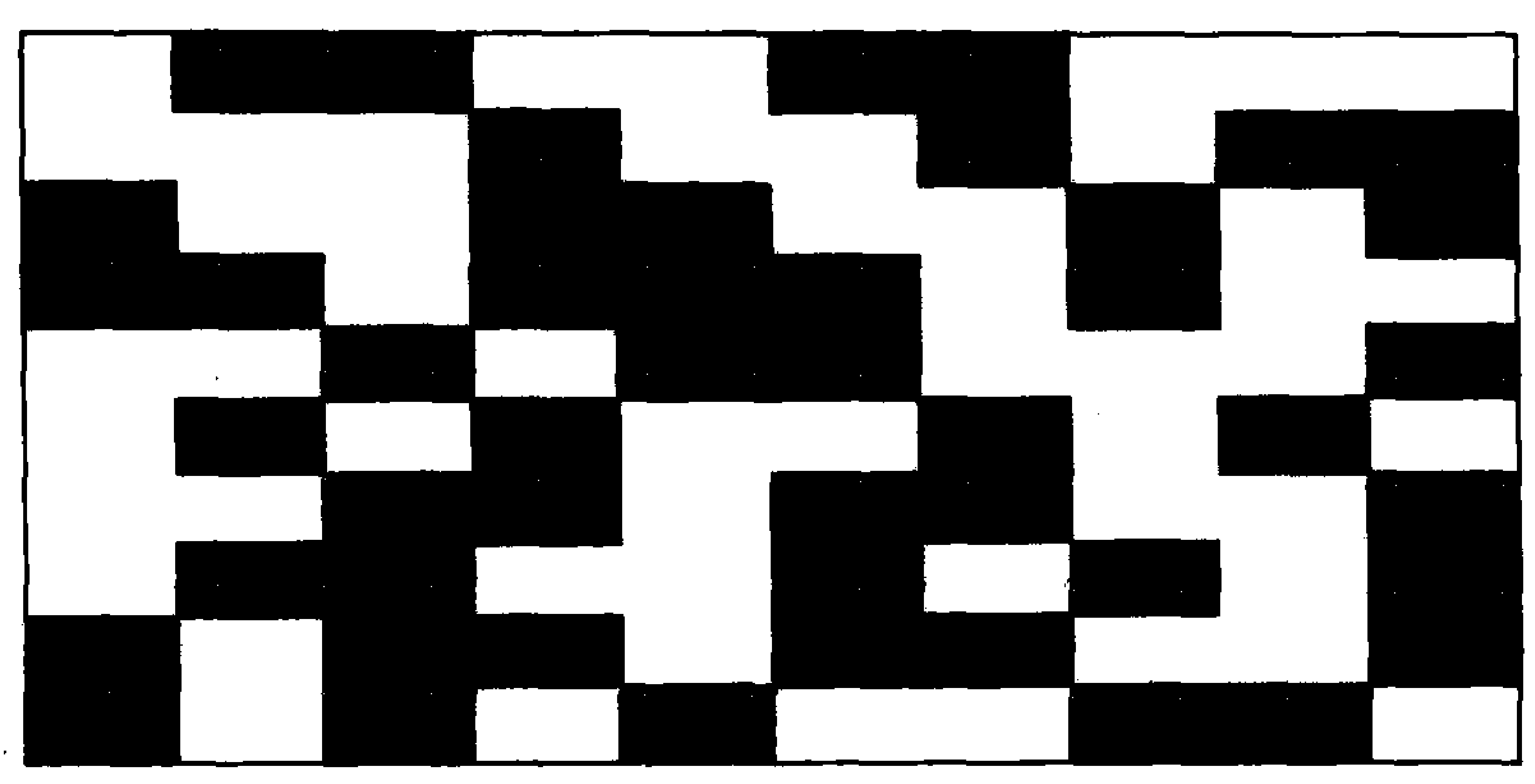


图 4.3　SCT 实验中对比敏感的幻灯片

在实验之前的指导是为了消除利益和权力可能引发的冲突。被试被确保实验任务是合作性的,而不是竞争性的。被试在实验前不会看到其他被试,也不会在实验中看到或听到他的合作伙伴。被试也会被确保实验之后不会遇到他们的合作伙伴。这些保证是为了消除被试的行为可能受到的,像阿施研究中的权力期待的影响。

实验者解释研究分成两个阶段,这样他就可以知道被试通过哪个方法作出更正确的判断。解释后,实验就开始了。被试将会就他们作出多少正确的判断而得分。一组被试(一个真实,一个虚拟)坐在分隔的房间中。每个阶段包含许多次测试。在第一个阶段的每次测试中,被试看幻灯片5秒,然后按下“黑色”(阴影)或“白色”按钮来记录他们的判断。在第二个阶段中,被试被给予更多的时间和信息,因为他们要看40张幻灯片。在第一阶段,每次看一张幻灯片,就作出判断(现在被称为**最初选择**(initial choice))。紧接着最初选择,每个被试会看到他的合作伙伴的选择,然后再看幻灯片5秒。然后,被试按下“黑色”或“白色”按钮来记录他的最终选择。

实验仪器有两个组件:第一个是幻灯片投影仪和所描述的幻灯片,第二个是实验者控制的信息。被试通过一个由乔瑟夫和他的同事(伯杰、科恩和泽尔底切)发明的**交互控制机器**(Interaction Control Machine,ICOM)过滤初始反应,获得信息。看到开关控制的灯一亮一暗,实验者可以看到被试最初的选择。然后,实验者按按钮,控制ICOM的面板,打开被试面板上的灯,显示虚拟同伴的选择。根据测试中实验者对被试的反馈显示,虚拟同伴在40次初选中28次**不同意**,剩下的12次同意。今天此类实验以同样的方式有效进行,但PC机取代了幻灯片投影仪和ICOM。PC机的设置在第5章会进行描述。

穆尔如何在这个情境中测量影响力呢?穆尔使用伯杰及其同事发明的方法测量**影响力的阻力**。被试如果没有改变他的判断,使他的报告与虚拟同伴一样,那么他就没有受到影响。例如,被试最初按的是黑色按钮,看见他的同伴按白色按钮,然后他还是按黑色按钮作为最终选择,那他在测试中就没有受到

同伴的影响。这种模式称为**停留反应**(stay response)。**只有通过改变判断使最初的不一致变为一致时才是影响。**穆尔使用ICOM记录停留和影响反应的实例。下一步,他使用如下公式计算被试停留反应的可能性:

$$p(S) = \frac{N_S}{N_D} \tag{2}$$

其中,N_S 是停留反应的频率,N_D 是测试中同伴最初不同意的次数,p(S)与影响力反相关。即人们的p(S)值越低,就越容易受到影响。

所有的被试都是一所大专的学生,其中一半被告知他们的合作同伴来自当地的高中,即一半的被试被告知他们的教育地位高于他们的虚拟同伴。另一半被告知他们的同伴来自当地大学,这样被试的教育地位就低于他们的同伴。这两组人再各分成两个小组。一个小组仅仅获得他们参与者地位的信息,剩下的小组被告知教育地位对于任务成功完成有积极作用。因此,穆尔创建了4个初始条件:

1. 被试被告知他的同伴是一个高中学生(被试地位高,HS)。

2. 被试被告知他的同伴是一个高中学生,大专学生通常做得比高中学生好(被试地位高,明确相关性,HS-ER)。

3. 被试被告知他的同伴是一个大学生(被试地位低,LS)。

4. 被试被告知他的同伴是一个大学生,大专生通常做得没有大学生好(被试地位低,明确相关性,LS-ER)。

理论预测地位高的人比地位低的人少受到影响,即他们有更高的p(S)值。这就是穆尔所发现的。条件1和2中地位高的被试停留反应的比例高于条件3和4中地位低的被试(HS = 0.696, LS = 0.613)。实验支持SCT的预测,将地位与影响联系起来了。此外,穆尔的实验同样显示了地位与影响之间的联系符合理论所表述的过程。

SCT断言,除非有相反的证据,否则人们会利用地位信息来推断任务的有效性。没有被告知任务相关性的被试应该会使用他们相对地位的信息来推断教育地位与任务相关。因此,

被告知相关性信息和未被告知相关性信息的两组间的 p(S) 值没有区别。理论得到了支持。穆尔发现地位高的被试无论有没有被告知相关性信息,都有几乎相同的 p(S) 值(HS = 0.699;HS-ER = 0.693)。同样,地位低的被试的 p(S) 值没有受到相关性信息明显影响(LS = 0.634;LS-ER = 0.593)。因此,穆尔的发现支持了 SCT 的预测,地位影响了影响力以及地位如何影响到影响力。

我们对穆尔的实验运用 5 条标准(像之前一样,标准用粗字显示)。

1. **验证源于理论的一个或多个模型。**穆尔设计两个阶段作出决定的模型。在第一个阶段,人们独立作出决定。在第二个阶段,人们看了另一个真实或虚拟同伴的决定(影响来源),然后作出第二个或最终决定。

2. **联系初始条件和终止条件,运用理论进行预测。**穆尔使用 SCT 理论并通过改变理论的两个元素来作出四个预测。第一,相对地位的信息影响了影响力阻力:①地位高的被试阻力大;②地位低的被试阻力小。第二,相关性(将被试的地位与任务相关)没有影响到影响力阻力:①地位高的被试无论是否有相关性(已知或未知),影响力阻力相同;②地位低的被试无论是否有相关性,影响力阻力相同。

3. **建立复本,设定初始条件,并观察终止条件。**被试观察图 4.3 所示的幻灯片,作出 40 次测试的最初决定和最终决定(12 次同意,28 次反对)。两种不同的方法为预测设置了条件:①被试被告知他们有一个地位较高(大专对高中)或地们较低(大专对四年制大学)的同伴;②被试不被告知(或被告知了)教育地位与任务能力的相关性(例如,高教育地位意味着高任务能力)。对于不一致的测试,研究者观察最初的决定和最终的决定,看被试是否改变了判断使之与虚拟同伴的判断一致。

4. **将预测与结果进行比较,确定理论是否得到支持。**实验结果符合预测。地位高的被试比地位低的被试有更高的 p(S) 值,表明他们受的影响比较小。同样,任务的相

关性对于地位高或地位低的被试都没有影响。

5. 可以非常自信地从理论中推断，实例在理论上非常类似于支持理论的实验——但是预测不仅仅受限于类似的情况。

符合第五条，对于扩展这些实验的确很重要。穆尔实验之后有一些复本，其中一些用于不同地位特征如空军排行（机长、空警以及第三级的飞行员）或武装部队资格测试（Armed Forces Qualifications Test，AFQT）的评分。研究者有信心预测这些研究的结果，因为他们的设计类似于穆尔的设计。但是，其他实验运用两个或两个以上特征（一个空军机长获得 AFQT 的高分对一个空警获得 AFQT 的低分）。当多种地位特征一致时，理论得到支持：即当两个特征对于一个人都很高时，对另一个人就会都很低（见（Berger，Cohen and Zelditch，1972）中对这些研究的描述）。但是，当涉及多个特征时就会发生一个重要的问题：如果地位特征不一致会怎么样？为完成这个范围拓展，理论需要预测两个或两个以上特征的**不一致分配**（例如，高和低对低和高）。休斯（hughes，1945）在讨论种族和职业不一致配对时（例如，一个黑人（低）医生（高）诊断一个白人（高）病人（低）），提前 10 年预想到了这个问题。

理论家构想了两个选择性假设——平衡假设与结合假设。平衡假设提出小组成员通过忽视或不考虑一些地位信息来“平衡”情境。例如黑人医生或许忽视种族地位，白人病人可能忽视职业地位。作为选择，结合假设认为小组成员会以一种方式结合信息，类似于取平均值或权衡信息的平均值。人们设计了实验在两种假设之间进行选择（Berger and Fisek，1970；Berger，Fisek and Crosbie，1970）。实验结果支持了结合假设，使再形成的理论包含了结合机制。在这个案例中，设计实验起初在理论范围之外，从而引起理论中的新发现，并最终通过实验得到了支持（Berger et al.，1977）。

最后我们来评价一下穆尔对于多次实验的运用。理论预测地位高的参与者会比地位低的同伴更拒绝影响力。但是被试没有与真正的同伴相互作用。因此，穆尔对被试进行了多次实验，担任高地位或低地位的角色，比较他们的结果来确定理论的预

测是否正确。不同实验的类似性是基于理论驱动型逻辑的作用和穆尔严格实验的控制,而不是基于差异实验的逻辑。

几何学实验:光学和交换

由于理论是科学的方法,当来自不同科学的理论有类似的结构时,它们的实验也会设计得类似。物理学中的几何光学理论和社会学中的基元理论就是两个这样的理论。这两个理论都有原理和定律:光学中最重要的原理是光是直线传播的,基元理论运用合理性原理;两个理论都非常依赖于图样去构建模型,并且它们的原理在决定简图如何绘制和解释方面起了非常重要的作用。事实上,这两个理论都包含了三个元素:原理、定律和简图,由此而产生预测,并设计实验进行验证。

原理必须在理论建立的框架内运用,而不能独立运用。例如,光直线传播的原理断言"光是沿直线传播的",但反射改变了光的方向,折射像引力一样使光变形,因此,将光直线传播原理独立运用或将它作为经验的普遍化是错误的。当它适当地运用于几何光学理论中的一个元素时,它是正确的。在那个理论中,原理表明,在混合的光学图中,理论家认为光束是呈直线传播的。然后反射原理和折射原理显示,在光束遇到镜面或棱镜时是如何改变方向的。

与之类似,基元理论的合理性原理认为"所有参与者都最大化他们所期望的获利",这也应该仅仅运用于理论所建立的框架内,而不是独立运用。独立运用就好像断言所有人都是理智的,其行动都是由自私心所驱动的,不会根据其他人的行动选择行动。这些都是错误的普遍化。但原理不是普遍化的,它是理论关注参与者如何将首选与行为联系起来的一个方面。理论没有排除利他的选择。更普遍的是,这些理论和其他理论一样,原理是不能断言世界的。它们可以与规律一起创建现象的模型,产生预测,设计实验。

应用几何光学的理论

几何光学的研究可以追溯到公元前约280年的欧几里德。欧几里德是第一个断言光是直线传播的人。但是像柏拉图之后的其他希腊人一样，欧几里德所认为的光的传播方向与实际相反。因为，他认为光是从眼睛出发传播到外界物体上的，这与我们今天所知道的正好相反。欧几里德在他的**凯特光学**（Catoptics）（我们下面会应用的原理）中也正确陈述了反射原理。古人在折射方面的研究没有获得成功：第一个尝试的是亚历山大的托勒密，但最终被斯内尔在1621年推翻。斯内尔原理在16年后被笛卡尔独立发现。事实上，这样的独立发现在科学上是很平常的。

我们现在来解释光学模型如何用于设计实验，但我们不说明由谁进行光学实验。反射原理像杠杆原理一样，是非常老的实验，已经不可能说出谁是第一个实验者了。在绝大多数情况下，一般是学生在物理学入门课上使用在此描述的设计验证该理论。

反射原理认为

$$i = r \tag{3}$$

即，入射角 i 等于反射角 r。原理产生的模型如图4.4所示。图4.4是一个来自平面镜上的反射光线的模型。从左开始，直线代表光线，以与法线成 i 夹角射向水平镜面。法线，以虚线显示，是一条想象的与镜面成90度角的直线。右面是与法线成 r 夹角的直线，代表来自镜面的反射光线。

图4.4的模型是验证反射原理预测的实验复本。实验的

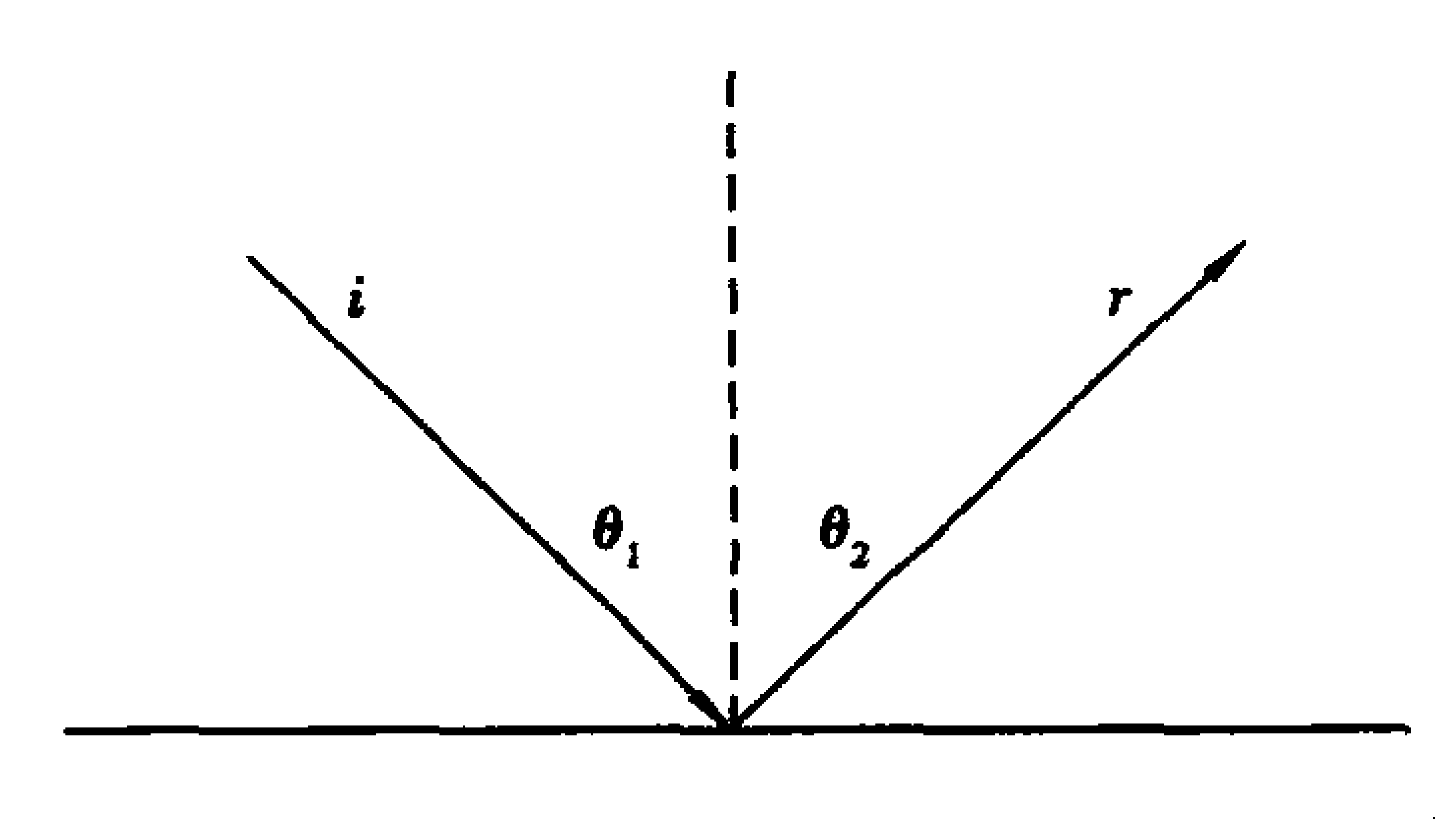


图4.4 平面反射

最初条件是一束光源、一个反光平面、角度 i。角度 r 可以应用原理来预测,预测可以与观察结果相比较。由于 i 可以是 $0° < i < 90°$ 中的任何值,r 也一样,模型就可以用于多次实验,只要改变 i,就可以观察对应的 r。想要进行该实验的人会发现激光点是最理想的光束源。

如图 4.5 所示,几何光学的理论通过简单延伸以上实验,发现了任意曲面上的焦点。用以下方式建立模型。首先,画一个曲面,然后在 A、B 两点画两条曲面的切线。每一条切线上画与切线成 90 度角的法线。在 A、B 处有两条平行光线从左面射向曲面。在 A 处,光线以角度 i 射向镜面,预测将以角度 r 反射。第二条光线入射在 B 处,同样也通过原理预测,$i = r$。镜面的焦点是两条光线的交叉点 F。

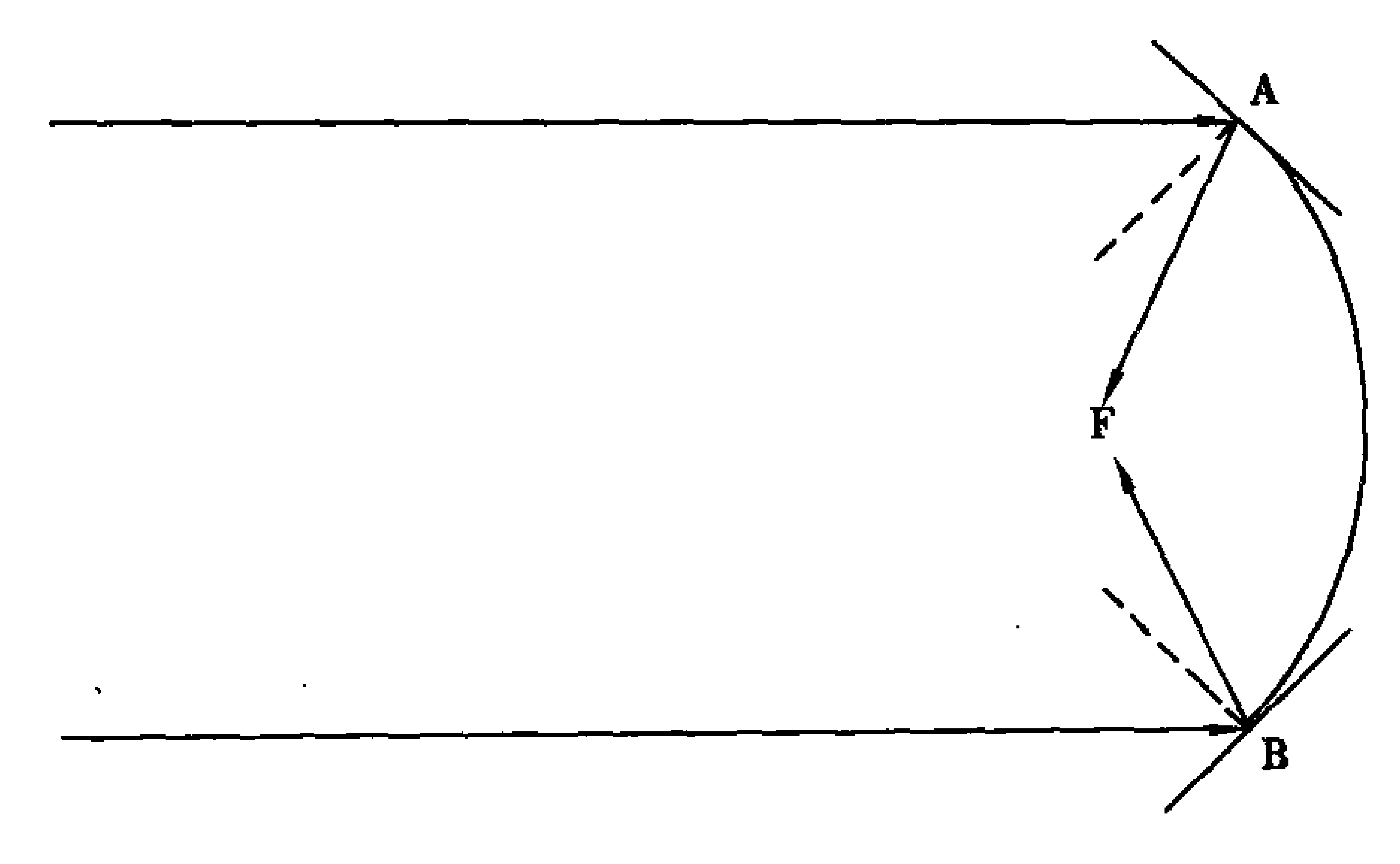


图 4.5 曲面反射

模型立即清楚地显示了如何建立实验复本来验证它。这个实验是之前实验的延伸,其初始条件是:创建两个相互平行的光束源、一个反射的曲面和角度 i_i。反射原理预测角度 r_i 和焦点 F,这些都可以通过观察进行比较(焦点是由镜子反射产生的可见影像)。再次设计模型进行一系列实验,在这些案例中,镜子的尺寸和弧线的半径都不同。

现在对上述光学设计应用 5 条标准。标准以**加粗字**表示:

1. 验证源于理论的一个或多个模型。上面提出了两个实验,一个是平面镜的,一个是曲面镜的。在每一个实例中,反射原理和光直线传播原理都产生了模型。

2. 联系初始条件和终止条件,运用理论进行预测。对

两个模型中的 i 都指定不同的值,反射原理产生对 r 的预测。

3. 建立复本,设定初始条件,并观察终止条件。 在这些实验中,建立平面和曲面的实验复本,设定 i——光线的入射角,观察结果 r——光线的反射角。

4. 将预测与结果进行比较,确定理论是否得到支持。 这些实验都进行了多次,结果无一例外地支持了理论。

5. 可以非常自信地从理论中推断,实例在理论上非常类似于支持理论的实验——但是预测不仅仅受限于类似的情况。

关于第五条标准,在讨论中已经包括了其重要的拓展:从平面镜拓展到曲面镜。增加了折射原理后,就会产生模型显示光束在通过不同介质(如,从空气通过玻璃棱镜再进入空气)后如何变形。通过结合比较简单的模型,可以建立比较复杂的模型;通过几个棱镜和镜面追踪光束,可以建模望远镜的各种设计——最终提供理论著述。法国数学家费尔马(Fermat [1662],1896),在1658年说明了光直线传播原理和反射、折射原理都符合光传播速度最快这个观点。

应用交换结构理论

至少从马克思(Marx,[1867] 1967)和韦伯(Weber,[1918] 1968)开始,人们就认识到交换可以使一方获利。交换关系是对称的:它们没有使一方比另一方多获利。它认为应该消除利益的不同,不是从交换关系本身,而是从其所处的结构和所置身的环境。为了解释结构中行使的权利,马克思构想出了一个概念"分离(separation)"。在资本主义的交换结构中,工人与生产方式分离,因为生产方式为资本家所拥有。分离加上储备工业的失业大军,意味着工人的工资会趋向最低,资本家的利润趋向最大。

韦伯同意分离是一个重要的结构性权力条件,因此使用它解释在机构中的权力集中化,因为在机构中,职员与他们所拥有的职位相分离。最近科拉(Corra,2005)将马克思和韦伯对于分离的定义解释为一种结构性的权利条件,与基元理论中最

近得到广泛调查的权利条件一样。那种权利条件是排除(Willer,1987,1999)。排除对于在交换关系之外的参与者比对于关系之中的参与者作用小。

这里所讨论的模型和实验调查了交换结构之外的影响。模型显示排除支持的参与者将会得到有利的交换比率和高获利,但以使他人得到不利的交换比率和低获利为代价。我们从显示如何建模一个交换关系,以及一个包含或不包含排除的结构开始实验。这里建立的模型与布伦南(Brennan,1981)建立的先驱交换实验中的模型非常相似。

图4.6的模型是包含交换关系的网络。每一个交换关系都包含两个带符号的矢量,称为反应(santion)。反应是相关的参与者发送或接受的行为,从中他们可以获利。讨论将关注A—B的关系,不过因为所有的关系都一样,因此也相当于描述了所有其他关系的条件。

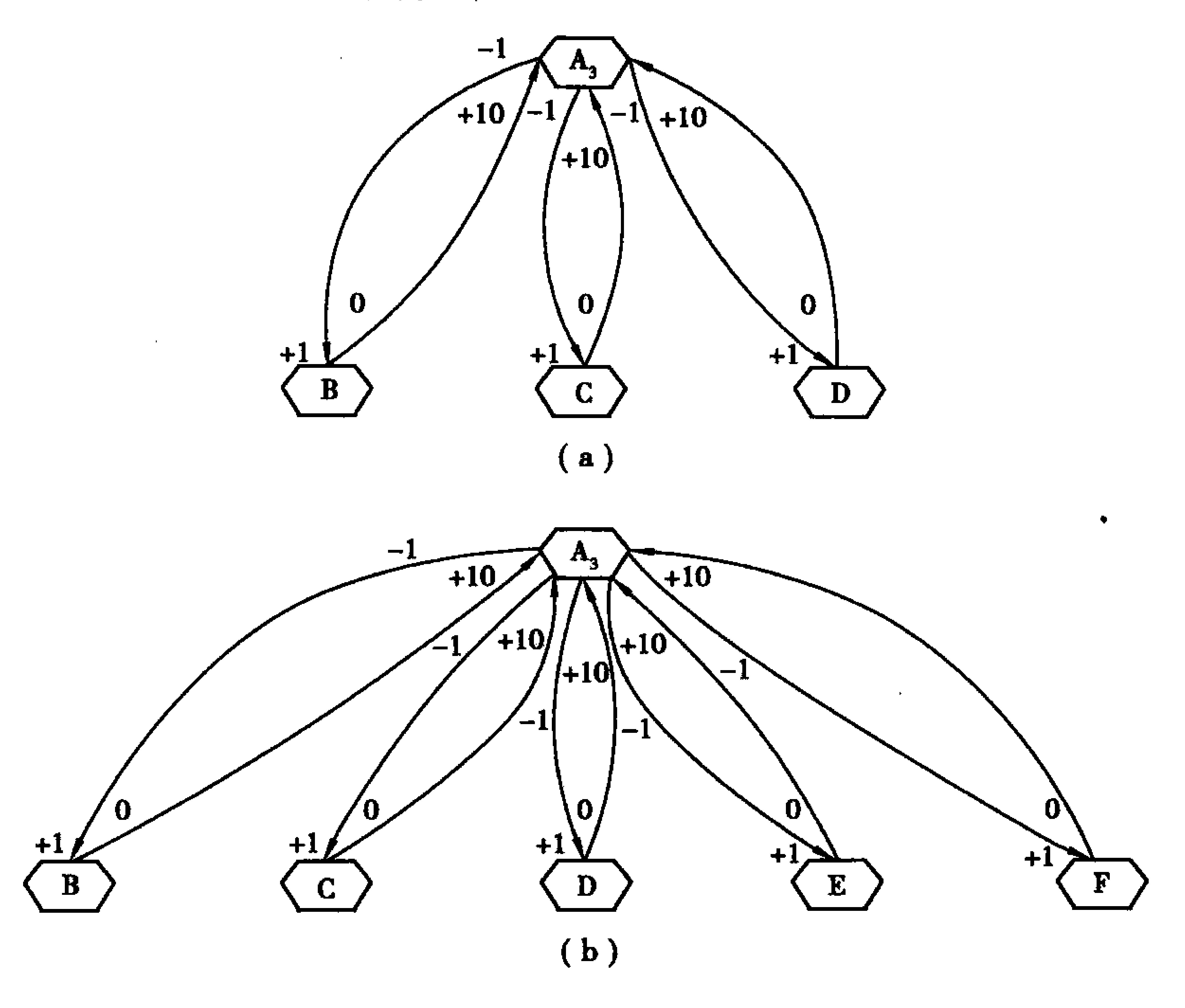


图4.6 无联系和专用联系交换结构

在A—B关系中(如图4.6a),B有一个积极的反应可以发

送给 A。给 A 的盈利是 10，表示为在 A 的反应矢量的末端"+10"。资源没有给 B 任何值，因此 B 的反应末端是 0。既然 B 持有一个反应，A 就可以发送一个或多个资源给 B（如符号所显示），每个反应都是对 A 的损失，对 B 的获利。A 的反应类似于金钱。A 为了获得 B 的反应而支付的反应数量为 x。事实上 A 与周围的三个参与者有相同的关系，A 的下标为 3 表示了 A 可以与全部三个周围参与者交换。

在什么情况下这些参与者会传递和接受反应，并且每个人都能获得盈利呢？从图中我们可以看到 P_A 是 A 的盈利，P_B 是 B 的盈利。$P_A = 10 - x$，$P_B = x$。即，A 的盈利等于从 B 的反应中得到的 10 减去发送给 B 的反应数目，B 的盈利只受到来自 A 的反应的数目的影响。A 与 C、D 之间有着同样的盈利关系。基元理论应用上述讨论的合理性原理。在 A—B 关系中运用该原理，我们可以确定他们的协商集合——理性的参与者交换的范围。协商集合包括所有 P_A 和 P_B 积极的交换。假设 A 的反应可以分到单位 1，当 A 收到来自 B 的反应，发送一个反应作为回报时：$P_A = 10 - 1 = 9$，$P_B = 1 - 0 = 1$。由于 P_B 仅仅收到 1，这就是协商集合的末端。在另一个极端，A 收到来自 B 的反应，并发送 9 个反应作为回报：$P_A = 10 - 9 = 1$，$P_B = 9 - 0 = 9$。发送 9 个反应是协商集合的另一端。从图表中可见，对于 A—C 和 A—D 的关系有着同样的交换范围。所有交换都在协商集合内发生，但是在集合中的哪里呢？除了合理性原理，还需要一些东西来预测交换范围中发生交换的点。

基元理论包括阻力原理，预测了图 4.6a 和图 4.6b 网络中发生的交换。$P_A max$ 是 A 获得的最大盈利，$P_A con$ 是在达不成协议时，受到冲突的 A 的盈利。阻力为

$$R_A = \frac{P_A max - P_A}{P_A - P_A con} \tag{4}$$

当公式(4)右面阻力因素显示了参与者的混合动机：①竞争获得更高的盈利，改变分子中的因素；②合作达成协议，改变分母中的因素。基元理论的第二个原理断言参与者的阻力因素相等时，他们会达成协议。A—B 的关系为

$$R_A = \frac{P_A max - P_A}{P_A - P_A con} = \frac{P_B max - P_B}{P_B - P_B con} = R_B \quad (5)$$

A 的最好盈利 $P_A max = 9$ 在协商集合的一端,B 的最好盈利 $P_B max = 9$ 在集合的另一端。当 A 和 B 无法达成协议(冲突)时,不会产生反应,谁都不会获利。因此冲突时的盈利,即 A 和 B 的 Pcon = 0。在阻力公式中替换这些值,$x = 5$,$P_A = 10 - 5 = 5$,$P_B = 5$。即,基元理论预测 A 会为了获得 B 的 1 个反应而交换 5 个反应。由于三个参与者与 A 的关系都相等,同样的预测也适用于 A 与 C、D 的交换。

$$R_A = \frac{9 - (10 - x)}{(10 - x) - 0} = \frac{9 - x}{x - 0} = R_B \quad (6)$$

图 4.6b 中的网络由与图 4.6a 中相同的关系构成,但现在有结构化的条件,排除影响着交换和盈利。A 仍然可以做三次交换,但现在可以从五个参与者中选择。正如上面所计算的 A 与 B、C、D 达成了协议(暂时)后,我们描述的这个结构的原动力。由于 A 有三个提供者,E、F 面对着排除,不与 A 进行交换,盈利为零。由于 A 是理性的,E 和 F 必须比 B、C 和 D 提供更好的给 A,就是从 A 处少要求一个资源;那么 $x = 4$。考虑 A—E 关系:如果 A 没有和 E 交易,A 会从与 B、C 或 D 达成的协议中获利 5。因此,A—E 关系中,冲突盈利 $P_A con = 5$。E 所能希望得到的最好盈利是 $x = 4$:$P_E max = 4$。

$$R_A = \frac{9 - (10 - x)}{(10 - x) - 5} = \frac{4 - x}{x - 0} = R_E \quad (7)$$

解阻力方程,$x = 2.5$。因此,$P_A = 7.5$,$P_E = 2.5$。F 会如 E 一样提供相同的资源,B、C 或 D 也会作出相同的交换。一旦 A 接受了 3 个 $x = 2.5$ 的资源,只有提供一个更好的资源给 A 才会避免被周围参与者排除。因此 A 会接受越来越好的资源,P_A 会增加直到 $P_A = P_A max = 9$。相应地,周围参与者的 P 值会减少到 1。因此三个周围参与者在另外两个被排除,获利零时,会相互交换得 1。

布伦南(Brennan,1981)使用图 4.6 中的网络模型设计了一对复本来验证排除是否会如图 4.6b 网络所预测的那样产生影响,没有排除时是否会如图 4.6a 所预测的有盈利,这些都可

进行观察。模型是为可重复的实验安排的，因此实验室的物理布局也要做相应安排。被试的座位按照相应字母安排。每个交换关系建立如下：在 A 的座位和其他周围参与者的座位之间放置一张小桌子，筹码堆积在每张桌子上。实验规则解释，B 有 10 个蓝色的筹码，以 1 为单位发送——其他周围参与者类似。规则还表明，筹码对周围参与者都无价值，而对于 A 每个筹码代表 1 个资源。A 有 33 个筹码。每个筹码对于 A 代表 1 个资源，对于发送的周围参与者而言也代表 1 个资源。在结构中，所有被试都被告知，A 最多可以进行 3 次协议，并只允许在已知关系中进行协商。周围参与者是被分隔物分开坐的，以消除他们之间可能进行的协商。

为了产生模型中的优先选择，被试：①在他们的交换中获得点数；②被要求尽可能地获得更多的点数；③被告知不要过分关心他人获得的点数。

布伦南以下面的方式进行了每项实验。每个结构以四个回合协商和交换开始。每一回合都由同在研究室的实验者计时。每一回合都持续 3 分 20 秒，被试在 3 分钟时会被提醒时间，以保证他们完成还在决定中的交换。当时间到了或者当所有交换都完成了，无论哪个先完成，实验者重新分配资源到他们初始的位置，开始下一回合。在四回合结束时，第一阶段结束。然后所有被试旋转：A 移到 B 的位置，B 移到 C 的位置，依此类推，F 移到中间成为 A。然后第二阶段开始，进行的过程与第一阶段一样。第二阶段后是第三阶段，一切都类似，直到所有被试都坐过所有位置。因此图 4.6a 网络的实验有 4 位被试和 4 个实验阶段，图 4.6b 网络实验有 6 位被试和 6 个阶段。注意，通过旋转，消除了协商技巧的个人差异、地位暗示或其他的特殊差异所带来的影响时，保护了结构。

图 4.6a 的网络中，A 从各交换对象处得到的盈利平均值是 3.1，提供给外围参与者的盈利平均值是 6.9。结果不太符合预测。在图 4.6b 的网络中，A 的盈利平均值是 7.9，每个交换的外围参与者平均获得 2.1。图 4.6b 网络的测试数据表明 A 的平均盈利小于 $P_A = P_A max = 9$ 是因为，虽然每回合盈利增加，但它们没有立即达到极端。事实上，A 的盈利模型预测

$P_A=9$。因此图4.6b网络实际上是符合预测的。回想实验的目的是为了证明排除的效果,从结果来看是成功了。当排除存在,A平均盈利7.9,当排除不存在,平均盈利3.1。因此排除存在比不存在多盈利2倍(对A而言)。

另一方面,图4.6a网络中,A—B的交换是3.1—6.9,远远偏离预测的5—5盈利。那就有两种可能性,可能是验证的理论不正确,也可能是实验没有精确地复制模型。如果理论是正确的,实验结果可以显示,被试是否得到了没有包含在模型内的反应。例如,外围的被试可能通过抱怨或侮辱对A产生消极反应,也可能临时对A反应,希望诱使A给出对被试更好的盈利。如果这样的话,他们可能使所观察的交换远离期望值。进一步检查发现,显然是外围的被试引进了模型中没有的反应。当他们获得低盈利时,他们抱怨中心的被试贪婪。

被试声称中心被试贪婪有没有物质基础?如果交换如预测的比率发生,中心和外围的被试都会在每次的交换中获得5个点数。然而在每回合结束时,外围的每个被试只获得了5个点数,即中心被试发送了5个反应。相比之下,中心被试从3个外围被试发送的反应中获得了10+10+10=30个点数,并获得了33-5-5-5=18个额外点数。33是最初留下的资源值。因此,按照预测的交换比例,中心被试在每一回合后获得30+18=48个点数,而每个外围的被试仅获得5个点数[9]。

一个修正分数的系统被创建出来用于消除外围被试引进的模型中本来没有的反应。新的系统反转了所有的交换关系,如:A最初拥有无价值的反应,B等拥有资源(诸如钱),保留着是有价值的。现在,如预测$x=5$时交换,中心A在每一回合结束时获得5+5+5=15个点数。每个外围的被试也获得15个点数:10点来自A发送的反应,加上5点交换中没有使用的资源。这个设计(1)消除了来自外围实验者的抱怨,(2)使平均交换没有明显地与理论预测的$x=5$不相同。这个修改支持了预测,同时并没有排除公平规范也对出现$x=5$的结果产生作

9 每一回合后宣布收入,正如我们指出的,外围被试会意识到不平等性,并对此做出反应。

用的可能性。

现在我们对布伦南的实验应用5条标准。同样标准以粗字体显示：

1. 验证源于理论的一个或多个模型。 两个模型都从理论及其原理中得出。第一个模型是没有排除的结构，而第二个是存在排除的结构。

2. 联系初始条件和终止条件，运用理论进行预测。 理论的阻力原理和阻力原则与合理性原理合在一起，就提出了两个模型的相反预测。它们预测了在没有排除的结构中的平均交换比例，和存在排除时的极端比例。

3. 建立复本，设定初始条件，并观察终止条件。 在实验室的布置中，实验复本就像模型一样安排。此外，设计实验盈利和实验要求，使被试在符合模型条件的情况下进行优先选择。被试面对着各自的资源筹码，可以在达成共同协议时进行交换。两个设计的不同只有一个初始条件：外围被试的任务(3个或5个)。由于A只允许每一回合与结构中的3人进行交换，5个外围被试的结构中就会有两人被排除。

4. 将预测与结果进行比较，确定理论是否得到支持。 图4.6b结构中存在排除的预测符合结果。图4.6a结构中的预测没有得到支持，但是重新应用理论，创建新的初始条件时，新的实验使预测得到了支持。

5. 可以非常自信地从理论中推断，实例在理论上非常类似于支持理论的实验——但是预测不仅仅受限于类似的情况。

关于第五条标准，在此描述的实验是基于基元理论的许多研究中第一次实验应用。当 $N > M$ 时，排除发生，其中 N 是可以交换关系的数目，M 是完成交换的最大数目。在图4.6b的结构中 $N_A = 5 > 3 = M_A$。排除只是基元理论研究者们及其实验调查发现的七种权利情况结构中的一种。一个**虚无联系**是 $N = M > Q = 1$，其中 Q 是一个位置上所有能从交换中获得的利益都完成时所必需的交换次数。由于 $N_A = 3 = 3 = M_A > Q_A = 1$，图4.6a网络就是**虚无联系**($Q$ 也等于图4.6结构中的排除联

系)。由于对所有的七种结构情况的描述和对之进行的实验调查随处可用(Walker et al. ,2000;Willer,1999;Corra and Willer,2002),在此不再对之进一步讨论。

由于基元理论的范围被拓展了,它的核心概念所剩无几。例如,对于所有的七种权利情况的预测产生了合理性原理、阻力定律和阻力原理。这些相同的理论观点用于预测强制结构(Willer,1987)和交换结构,资源可以在其中通过网络分多步流动(Willer,2003)。这些拓展都表明了社会结构的研究只是刚刚开始,未来还有很多需要去研究。幸运的是,网站上(http://weblab. ship. edu)免费提供用于建立实验的软件。它可以用于重复布伦南的交换实验。软件及一些关于软件的使用将在第6章中进行讨论。

总　结

本章在开始时指出,社会学中的理论驱动型实验与其他科学的实验在方法论上是相同的。它们在这一点上是相同的:理论设计实验。这里提供两个证据。第一,本章介绍了三对实验,每一对实验都是一个来自物理学,一个来自社会学。倾向于经验性的研究者肯定会发现,尽管每对实验的领域完全不同,但比那些来自同一领域的实验有更多的相似性。有一个例外,即第二个实验与最后讨论的实验都运用基元理论。因此,那两个彼此类似。

对于倾向于理论性的研究者,我们提出了第二个更强有力的证据。理论驱动型实验的原理在本章的前面部分已提到,并作为标准应用于所讨论的全部(六个)实验。这些标准对于物理学和社会学实验同样都能很好地符合。两者的根本观点是建立类似源于理论模型的实验复本。然后实验验证模型与复本这二者的相似性:从初始条件到终止条件。

社会学中的理论驱动型实验与其他科学的实验相比并不特别,也不比它们差,因此是时候撇开以下观念了:因为实验结果不能普遍化,因而实验的应用是有限的。根本不必从理论驱

动型实验中得出一般结论。相反，正确的科学过程是将实验验证过的理论应用到实验室外的研究个案中。

例如，伊丽莎白·G. 科恩，一位首创 SCT 的研究者，在实验室外应用了 SCT。科恩认为阅读能力是一种地位特征（Rosenholtz and Cohen，1983），也同样意识到学生在阅读能力方面存在差异。运用 SCT，科恩预测其他学生和老师会使用阅读能力（就如同他们使用其他特征，像种族或性别）来建立对学生智力技能的期望。此外，他们还使用某种特征（如种族或性别）来建立对阅读能力的期望。相应的，学生和老师基于学生实际的或被认定的阅读能力与学生相互作用，作用的方式是试图证明关于学生在学校表现如何的自我实现预测。因此，低地位的孩子阅读能力逐渐变得没有高地位的孩子好。科恩没有就此止步，运用 SCT 设计了教室中的干预策略，扰乱这些过程，帮助低地位的学生更好地学习（Cohen，1998；Cohen，Lotan，Scarloss and Arellano，1999）。

科恩提出了完整的地位和影响循环圈的研究过程。它从问题开始，以理论形式设计问题的答案，使用类似穆尔的实验验证理论，最后显示实验支持的理论可以解决实验室外的问题。现在，在第 5 章我们会转而探讨有关建立实验的实际问题。第 6 章会更多地讨论通过理论联系实验和非实验研究，包括基元理论在实验室外的应用。

5 实验中的社会关系

在社会科学和行为科学中，实验设计创设了包含社会关系的社会情境。正如马丁·奥恩（Orne，1962：277）所述，在那些关系中“被试和实验者的角色完全得到理解，并由此带来了明确的相互角色期待”。实验者—被试的关系决定于由文化和制度定义的规则和标准，而规则和标准同时也指导和限制了角色的行为。进而，实验者—被试关系影响了研究过程的每个阶段，社会学家必须考虑关系的社会角色是如何影响研究过程的。

在实验者—被试关系中，实验者与被试相比，通常有较高的地位、较大的权力，并更有权威。这样的不平等对被试和实验的影响都很大。实验者可能会滥用他们的权力和权威来压迫被试，也可能会影响被试让他们去做在研究环境外不愿意做的事情。我们在本章的第一节就开始关注这个问题，讨论道德标准，即明确人类被试的权利以及研究者的职责的科学标准。

招募被试建立了实验者—被试关系。招募被试很值得研究，因为招募所需的时间与精力不比进行实验少。我们进行了两项招募研究，从如何招募被试的具体细节，到有关被试报酬的道德问题。

实验者在实验情境中更大的影响力和权力也会潜移默化地影响实验研究的结果。这些影响无论与指导研究的假设或理论是否一致，都会被作为人为因素、需求特性以及实验者偏差被讨论。这样的讨论发现了研究环境、实验者和被试对环境的反应以及实验者偏差如何影响实验结果。

本章在总结时介绍了以计算机为媒介的现代实验情境。这些情境旨在有效地控制社会学实验中的社会关系，并且通常情况下能够发挥作用。在希望复本尽可能相似的情况下，这些情境还提高了理论模型和实验复本的匹配程度。用计算机界面进行交互，实验者偏差和被试偏差都得到了减少。事实上，提高匹配程度、创建精确的复本以及调解实验者与被试的关系都能减少本章中所探讨的各种类型的人为因素，包括需求特性和实验者偏差。幸运的是，感兴趣的学者可以免费使用我们所讨论的两个研究案例所用的软件。

“第一，没有伤害”：对人类被试的道德及其研究

物理学家在粒子屏幕上击碎亚原子粒子后，或许只有喜剧演员才会对该粒子身体或情感上的健康表现出关心。相比之下，医学、社会学和行为研究者也可能会对他们研究的人造成长时间的生理、心理和情感上的伤害。因为有造成伤害的可能，所以他们有责任确保不会危及人类被试的健康。医学教授的格言“第一，没有伤害”，是所有研究者对待人类被试的正确规则[1]。正如我们现在所说明的，在社会学和医学中，许多研究没有遵守这条规则。

两项研究

塔斯基吉梅毒研究 1932 年，美国国民健康服务体系在乔治亚州梅肯郡开展了一项研究。塔斯基吉梅毒研究涉及 600 名黑人男性，其中 399 名在研究开始时就已患上梅毒，而另外

1 这句格言与希波克拉底誓言(Hippocratic Oath)有关，但不是誓言的一部分。而且，“不以伤害为先(primumnon nocere)”既不源于拉丁语，也不是拉丁语短语的本意。根据史密斯(Smith,2005)所述，医学历史学家们在《希波克拉底流行病》(第一卷书，第六章，[公元前 400 年] 1932)中发现了类似说法。但是，希波克拉底或盖伦(古罗马的医师，另一个假定的作者)不可能是格言的出处，因为这些都早已记载在希波克拉底的出生地希腊。实际上，格言的出处远没有格言本身所传达的信息重要：那些涉及人类被试的研究哪怕付出一切代价也要保护人类被试的健康与安全。

的201名没有患上梅毒。那时此病没有可靠的治疗方法,没有受过教育的穷人大多受到感染。当时最好的治疗方法是使用毒性很强的药物,而治愈概率一直维持在30%左右。国民健康服务体系向研究被试承诺,会在他们的疾病治疗期间,对他们的健康状况进行监测。

塔斯基吉梅毒研究实际上是一个差异法的实验,旨在观察**未治疗**的梅毒对健康的长期影响。研究者将受感染的被试与未受感染的被试的健康状况进行了对比。20世纪40年代后期,使用青霉素成为该疾病的标准治疗方法,但在塔斯基吉研究中没有给被试使用这种药。而且,无论是被试还是普通大众都没有意识到研究的本质,直到《纽约时报》在1972年7月曝出内幕——距离研究开始已经40年了(Jones,1981)[2]。

斯坦福监狱实验 1971年8月中旬,在塔斯基吉研究内幕曝光一年前,美国加利福尼亚州帕洛阿尔托的警员拘禁了9名暑期班学生。所有学生都是男性,他们的"拘禁"是斯坦福实验(Stanford Prison Experiment,SPE)的一部分,该实验由心理学家菲利普·津巴多负责(Haney,Banks and Zimbardo,1973)[3]。据津巴多的记述,他和他的研究伙伴们想了解"当我们在匿名环境下,将所有这些过程汇集在一起时会发生什么——使一些参与者感到被剥夺个性,另一些参与者感到失去人性,这些构成了受控实验环境下的'完全环境'"(Zimbardo,Maslach and Haney,2000:9)。

津巴多和他的同事从70名学生志愿者中进行甄选,最后剩下24个表面看上去正常的男性。随机选择其中的一半充当囚犯,另一半充当看守,在斯坦福大学心理学系大楼地下室临时建立的监狱内进行实验。津巴多和他的同事在**可控环境**(controlled environment)下进行他们的研究。研究者没有控制情境特征,诸如剥夺个性或失去人性(Zimbardo et al.,2000:9),而

2 在这场悲剧中,美国政府的角色是不道德的。但是我们必须指出,许多医疗从业者和医疗机构,包括美国医学会和有影响的黑人全国医学会至少在20世纪60年代末就察觉到了这项研究存在的问题,并容忍了它。

3 研究假设见 http://www.zimbardo.com。作为改编,可能不完全忠实于原著,实验可见于DVD。

且不能确定被检验的理论或理论模型。因此,它既不是差异法实验也不是理论驱动型实验。

研究者在仅仅6天后就停止了SPE,虽然它原本计划持续2周。在这6天内,囚犯反抗,看守变得充满恶意而又残忍。当时,看守打破了他们之前建立的对待囚犯的规则(例如,对囚犯单独关禁闭的时间超过了规则所允许的时间)。一些家长进行了干预,并诉诸法律顾问将他们的孩子解救出"监狱"。此外,津巴多和其他"角色扮演者"的表现也不像研究人员在探求未知,而更像是采取高压手段的监狱当局平息囚犯之间的动乱。

塔斯基吉研究和斯坦福监狱实验违反了**不以伤害为先**的原则。许多塔斯基吉志愿者在可以治疗的情况下不必要地忍受梅毒的病痛,因为得不到治疗,许多人过早或痛苦地死去。SPE中的一些人承受了严重的心理和情感创伤,5个人在研究者结束实验前就退出了。这些研究和无数其他类似的研究都提出了棘手的道德问题:研究者对人类被试的责任是什么? 人类被试有什么样的权利?

人类被试的权利

每一位研究者在进行涉及人类被试的研究时,都有责任保护他们。科学界为参与医学和行为研究的人类被试建立了一部《权利法案》。美国国家卫生研究院的贝尔蒙特报告(Belmont Report)建立了一个宽泛的框架,概述了人类被试的权利以及研究者对他们的责任(National Institutes of Health, 1979)。他们的权利包括但不仅限于以下这些方面:

1. **被试有权免受不必要的风险**。这里的风险包括情感上、心理上或生理上潜在的伤害。

2. **被试有权完全了解研究过程**。

3. **被试有权完全了解任何和所有潜在风险**。研究者必须保护被试免受伤害,他们也必须告诉被试研究过程中任何已知的风险,无论风险多小。向被试告知所有已知风险,有助于被试有充分依据决定是否开始或继续参与研究。

4. **被试有权拒绝参与研究项目**。一些学术研究者要求学生参与一个或更多个研究项目以换取部分课程的学分。但是,

学生正如所有其他人类被试一样,绝对有权拒绝参与研究项目。

5. 被试有权在任何时间退出研究项目。研究被试可以同意参与研究,也可以在研究结束前提前退出。如果被试决定提前退出,不得强迫其继续。

6. 被试有权免受强制参与。无论从心理上还是生理上,都不能迫使被试不知不觉地参与研究。

实施道德规范 联邦法规和研究机构自身的惯例都要求研究者:①有对人类被试做研究的证明;②有经核准的在研究中保护被试的方案。作为认证过程的一部分,要求研究者证明自己熟悉被试的权利、研究者对被试的责任,熟悉保护被试权利的流程。许多机构,包括我们所在的学校,都需要研究者通过由罗彻斯特大学指定材料的考试(Dunn and Chadwick,2001,有时称为"罗彻斯特协议")。

关注人类被试并不限于对研究活动的联邦监督,人文社会科学的多个专业组织在各自领域都有规范研究者行为的详尽的道德守则。下面我们就将提到《美国社会学协会道德守则》(1999)的部分内容[4]。

设计和开展道德研究

研究者计划进行涉及人类被试的研究时,应该联系他们当地的机构审查委员会(IRB),咨询保护人类被试的流程。多数IRB都是一些机构(例如,研究性大学或医疗中心)的组织单位,这些机构定期参与人类被试的相关研究。也有一些IRB是独立组织,它们审查项目的范围包括:独立承担项目的研究者,或者因为规模小而没有自己的审查委员会的研究机构。研究者,包括研究助理,都必须在开展研究前进行培训或测试准备工作,确保安全认证完好。通常情况下,申请政府支持或基金会支持的

---

4 美国社会学协会的道德守则(1999,http://www.asanet.org),美国人类学协会的道德守则(1998, http://www.aaanet.org),以及美国心理学协会的道德守则(2002, http://www.apa.org)——括号内的地址都指向它们相关站点的首页链接。这些协会会定期更新它们的道德守则。

研究者在资金注入他们的项目之前必须提供相关证书。

所有研究项目都必须策划保护人类被试权力的方案，并提交给一个 IRB[5]。提交给 IRB 的研究方案必须详细描述计划开展的研究将如何保护被试的权利，也必须说明研究者如何检验被试在开始参与研究之前已经了解他们的权利。必须在进行研究前确保获得 IRB 的批准。研究者也必须将反对项目的意见告知 IRB 或其中的一位办公人员。如果没有遵守这些道德规则，将受到严厉惩处，包括罚款、收回研究资金或失去申请研究资金的资格。

实践中的道德标准：社会学研究中的欺骗

塔斯基吉实验中的被试不知道他们的疾病当时没有进行治疗，或者不知道 1947 年以后有了治疗梅毒的有效药物。穆尔实验（第 4 章）的被试不知道他们是与机器“互动”。这些研究似乎都违反了被试“权利法案”的多项规定。首先，在这些研究中的被试都没有获得关于研究流程的正确、完整的信息，因此，他们同意参与并不是完全知情的。第二，塔斯基吉研究没有告知被试参与实验的潜在风险。此外，塔斯基吉实验和斯坦福监狱实验的研究者没有保护他们的被试免受生理上和情感上的伤害。参与两项研究的被试都遭受了暂时或永久的伤害，而这些风险本可以（或应该）预见得到。然而，没有证据表明，在穆尔研究中被试以任何方式受到伤害。

类似穆尔实验中的欺骗是合法的吗？许多研究者在他们的实验中使用欺骗和误导，但其他研究者坚决反对这样的行为（Geller，1982；Kelman，1967）。欺骗引起了难解的道德问题，它似乎违反了完全知情的要求（但是看看下文的内容），它对被试隐瞒了潜在的风险。但是，无论阿施还是詹姆斯·穆尔在开始他们的研究时，都没有告诉被试他们将获得实验者或他们同伴的错误反馈。如果告诉被试会有错误的反馈将不利于达到实验的目的，因为被试不会像以前一样认真完成实验。那么研

5 有一些研究可以“豁免”于 IRB 的审查。但通常，必须由 IRB 审核是否属于豁免情况。

究者如何在公开的研究流程中遵守道德标准(包括不欺骗)但同时又保护他们实验的真实性呢?

多家重要社会科学专业协会(见注释4)的行为守则规定了什么情况下的欺骗合乎或不合乎道德规范。由美国社会学协会(ASA)建立的标准规定,在满足3个条件时,允许研究者在实验中有欺骗行为:

> 可以确定①"欺骗"不会伤害研究参与者;②"欺骗"是从研究的科学角度、教育角度或应用价值角度来判断的;③**如果不使用欺骗,找不到同样有效的替代方法**。(ASA,1999:14,section 12.05a,此处添加了强调)

当研究满足这些标准时,研究者可以使用欺骗,而且大多数都在实验最后,在研究的关键部分完成后向被试说明情况。穆尔遵循了ASA政策的要求。该政策指出,必须在**不迟于研究结束前**揭密欺骗行为(ASA,1999:14,section 12.05c)。尽管如此,研究者还必须仔细权衡欺骗性实验的潜在危险和利益。

欺骗的类型 心理学家和其他社会学和行为学家都曾研究过人们对欺骗的反应。他们的研究表明,欺骗和事后说明欺骗都可能对**被试和研究人员**造成情感上或心理上的伤害。欺骗的研究还包括评价用于逆转或减少欺骗潜在负面影响的技术。西伯界定了研究中使用的三种类型的欺骗:**隐式欺骗**(implicit deception)、**技术欺骗**(technical deception)和**角色欺骗**(role deception)(引自Geller,1982;也见于Sieber,1992)。

发生隐式欺骗时,被试不知道他们在被人研究。塔斯基吉梅毒研究就是一个声名狼藉的隐式欺骗案例(Jones,1981)。社会科学中也有隐式欺骗的不佳案例:劳德·汉弗莱斯在公共厕所看到男性同性恋相聚时假装看门人(Humphreys,1970),他还记录了他们的车牌号码,并从机动车辆管理局获悉他们的住址。那些人一直都不知道汉弗莱斯是一个研究者,直到事情发生大约一年以后,他出现在他们的住所,询问他们的个人和家庭生活情况[6]。

---

6 其实,在汉弗莱斯开始继续调查他在许多个月前观察的男性之前,改变外貌并没有什么意义,当然,他的被试没有选择。

在其他隐形欺骗中，很有可能一些“参与者”到死都不知道他们曾经参与过研究。例如，皮列文、罗丹和皮列文（Piliavin, Rodan and Piliavin, 1969）观察了他们的同伴在纽约地铁上假装争吵时旁观者的反应。这些实验的参与者只有在阅读科学杂志上的文章或碰巧读到关于**旁观者干预**（bystander intervention）研究的通俗性报告时，才可能知道这项研究。在公共场合进行的研究（类似于皮列文等的）如果对被试而言没有或很少有风险，就不需要获得批准（ASA, 1999: 12, section 12.01c）。

研究者掩饰设备或流程的作用，就是技术欺骗。如第4章中所述，穆尔告诉SCT实验中的被试，交互控制器（ICOM）上的灯会记录他们自己和他们同伴的最初选择。事实上，实验者控制了ICOM以确保70%的成对选择不一致。另一例实验室研究的技术欺骗案例是，米尔格拉姆宣称他的电击发生器向学习者发送了电击。

角色欺骗发生时，研究者隐瞒研究中其他人的身份。在阿施最初的研究中，有7名“被试”是实验助手，他们的角色是诱导被试给出错误答案。在米尔格拉姆的服从研究中，“学习者”一直是实验助手。

欺骗的有害影响　欺骗对被试和研究者都会造成心理伤害。想象研究者刚刚告诉你，你是骗局的受害者，你持续45分钟向其发消息的同伴并不存在。或者，你刚刚知道其他被试根本不是被试，而是为实验者工作的，目的只是为了误导你。你会哑口无言吗？会尴尬窘迫吗？会蒙羞受辱吗？会愤怒无比吗？任何人在得知自己被骗后，都可能经历其中一种或所有这些感受。被试被实验者欺骗后，最终往往产生深刻、持续的影响。

一些欺骗甚至使得被试做出残忍、不寻常或与他们价值观不符的行为。米尔格拉姆的服从实验是这些研究中的经典案例，类似津巴多的实验，这两人是高中时的同学（Zimbardo et al., 2000）。我们大多数人曾经做过（或者没完成）某些事，让自己感到抬不起头。有时我们会再次思考这样的事。如果在深夜听到尖叫声，我会拨打911吗？如果我看到发生严重交通事故的现场，却没有停车帮助，我是什么样的人呢？很多情况下，我们为

自己的行为辩护,告诉自己因为发生得太快,"我没有时间思考"。在其他情况下,我们假定其他人会去处理问题。心理学家把后一种反应称为**责任扩散**(diffusion of responsibility)。

欺骗产生的危害不仅限于知道研究者曾经欺骗过自己的被试,研究助手也会有类似的反应。毕竟,什么样的人能够坚持每天都要欺骗别人的工作,或更糟糕的,眼睁睁看着正常的大学生在模拟的监狱中沦为禽兽[7]?

职业道德很明确,大多数研究者也同意,在任何可能的情况下实验者都应该避免建立和使用欺骗被试的实验设计(Kelman,1967;Geller,1982)。但是,一些调查者却不得不欺骗。例如,医学研究者在一系列临床试验中会给一些被试服用安慰剂。如果不服用安慰剂,研究者就不能够断定健康状况的改变是真正的药物效果还是研究参与者的社会心理反应。但是,从道德和伦理角度,使用误导或欺骗的研究者有责任确保欺骗带来的负面反应(如果有的话)只是暂时的。他们有义务有责任确保被试和研究助手不会由于研究经历受到情感上的伤害。事后进行情况说明能够帮助研究者达到这个目的。

情况说明

> 许多大学生认为,心理学家在大多数实验中故意欺骗他们。如果大学生们对经济学家也抱有同样的看法,那么我们将无法控制实验。正因为这一原因,现代实验经济学家一直非常小心地在他们的所有实验中维护完全诚信的声誉。这样实验的成本就更高了。(Ledyard,1995:134)

我们并不同意莱迪亚德的观点,因为他误解了欺骗何时可以使用。道德标准很清晰。问题并不在于实验的成本,而在于不用欺骗是否就不能进行实验。而只有出现这种情况,欺骗才

7 必须重点指出,大多数社会学和行为学研究者的行为都是合乎道德的。例如,当时的斯坦福心理学研究生,克里斯蒂娜·马斯拉奇要求研究者停止斯坦福监狱实验。如今,作为加州大学伯克利分校的教授,马斯拉奇已成为研究工作倦怠的非人性影响的知名专家。

可以使用。此外，使用欺骗并不一定会失去对实验的控制[8]。一个良好的情况说明会使被试在实验过程中成为一个合作者。作为合作的一部分，被试可能会被要求暗中保持欺骗——通常情况下他们会这么做。

实验后的情况说明有两个主要目的：第一，确保研究者能够让被试充分了解自己参与过的实验；第二，它使研究者有机会将参与者的心理状态和情感状态恢复到刚进入研究时的状态。

研究后的情况说明应该是所有涉及人类被试欺骗研究的不可或缺的一部分。研究者必须为受骗的被试**释骗**(dehoax)，并使之**脱敏**(desensitize)(Holmes,1976a,1976b)。释骗针对的是实验者的行为(如欺骗的程序)。脱敏针对的是被试的行为，包括由欺骗或实验者公布欺骗行为引起的新认知或因此改变的认知。哪一个任务都不简单。我们以达夫·科恩及其同事的研究为例。

科恩、尼斯比特、鲍德尔和施瓦茨(Cohen,Nisbett,Bowdle and Schwarz,1996)设计了一系列的求异法实验：①了解是否存在南部"荣誉文化"；②确定荣誉文化是否与男性的攻击性有关。科恩等人以密歇根大学的男生为研究被试，以被试的出生地(南部和非南部)和侮辱性(具有侮辱性和不具有侮辱性)作为感兴趣的实验变量，做了一个 2×2 的实验设计。研究者假设，**如果存在南部的荣誉文化**，那么面对侮辱，南部的男性比其他地区的男性更具攻击性。

实验中的被试首先接触的是实验助理，实验助理向他们简单介绍了实验，并要求他们完成一份简短的人口统计学问卷。在完成问卷后，要求被试把问卷放在一个狭长走廊尽头的桌子上。当被试走在走廊上时，一个男性助手站在打开的文件柜边，"砰"地关上文件柜，转身，走向被试。助手在经过被试身边

8 我们引用莱迪亚德的评论是为了说明，如果被试认为有可能受骗就会将人为因素带入研究。人为因素会影响结果，并使实验结果难以或无法解释。我们在本章后面的章节中会讨论如何减少或消除人为因素，以及如何测试人为因素。这里，我们主要关注研究者如何利用良好的情况说明技巧来揭示欺骗，弥补负面影响。

时猛撞被试的肩膀,并骂被试是混蛋。另外两个助手(一个女性,一个男性),目击事件,并记录下被试的语言和非语言反应,包括面部表情。在事件发生后,每个被试都要完成几项任务,包括写下两个场景的结局。在事后情况说明阶段,被试会接受释骗和脱敏。

释骗 在情况说明阶段,研究者应该首先对被试进行释骗,说明事情真相。在释骗过程中,研究者告诉被试他们所使用的欺骗及其进行欺骗的原因。在这个阶段,实验人员只有使被试相信他们的诚实,释骗才能奏效。毕竟,他们刚刚揭示了自己的不诚信。

科恩等人开始事后说明时,先与被试讨论了研究设计。当研究者描述了各种欺骗(如三个助手的角色)以及为什么欺骗对实验成败很重要后,接下来,会让被试有机会讨论实验的真实目的。在那时,被试知道他们刚刚被骗,他们如何被骗,以及研究者希望他们在欺骗中如何反应——谎言被充分披露出来。科恩等人的情况说明工作还包括让有侮辱性行为的助手参与其中,并让他们与被试交流。

脱敏 脱敏是释骗的下一步。欺骗性的研究活动或调查者揭露欺骗会使被试重新审视和改变他们的自我定位。如果不加以纠正,在被试离开实验室后,被改变的自我概念仍会对他造成影响。情况说明者的责任就是使被试恢复到他们实验前的状态。科恩和他的助手有被试作答事后阶段问卷的结果。他们了解哪些被试在初步的情况说明和释骗[9]后仍然很愤怒。有经验的调查者会继续进行事后情况说明,直到所有的问题都得到对被试而言满意的解答。有道德的调查者在释骗和脱敏后,还会告诉那些仍然心烦意乱的被试,在离开实验室后他们可以在哪里发泄、抱怨,以及他们可以对谁倾诉他们心灵上和情感上的不适。

---

9 科恩等(Cohen et al. ,1996)进行过三项研究,在第二个研究的情况说明后对被试进行了调查。他们报告说,89%的被试在他们有机会与攻击他们的实验助手见面后,不再感到愤怒。其余仍感到愤怒的被试,他们的愤怒水平也没有高过量表的中间值。这个量表的尺度是从0(无愤怒)到7(极端愤怒)。

情况说明程序应当在实验者将所有被试带入实验情境之前建立好。在机构审查委员会(IRB)核准所有程序后,才允许研究者将人类被试带入实验。正如讨论所示,研究者付出很多的思考、时间和精力设计实验,以确保合乎道德地对待被试,另外,如果实验中使用欺骗方式,还要开发和使用包括释骗和脱敏的情况说明程序。

单单是道德原因应该就足以促使研究者避免使他们研究的被试冒风险。幸运的是,人们也在呼吁将"合乎道德地对待被试"落实到现实中来。我们知道没有一项研究表明,不道德的研究会产生与严格遵守道德标准的研究一样有用的科学发现。像塔斯克吉研究或斯坦福监狱实验那样的不道德的研究就是伪科学的例子。

招募被试

被试从哪里而来?研究者如何与他们获得联系?很明显,被试不会偶然地来到实验室,他们需要提供某些激励。然而,一种激励方式——作为课程要求,要学生参与实验——虽然已使用了很长时间,但如今已不再适用。时至今日,这样的要求被视为强制性的,因此也是不道德的——大多数IRB都不允许这么做。另一方面,把参与实验作为获得额外学分的一项选择却是合乎道德的。引导潜在的被试到我们的实验中的另一种方式是付给他们酬劳。虽然给被试酬劳对他们而言是一种好处,但这种做法存在缺陷。在下文中,我们将回顾一些酬劳带来的缺陷。

在招募被试和安排日程上所花费的时间和精力与进行实验不相上下。以我们的工作为例,比较好的情况下,一周内我们可以试验12个阶段的72位被试。每周全部实验的时间大约是30个小时,招募和安排日程所花费的时间总共超过了18个小时。我们现在来回顾一下曾经用于招募和安排大学生被试的一些流程。

尽管可以使用在学生报刊上刊登广告的招募方法,但是我们更倾向于在大班中进行招募。幸运的是,许多不同学科的教授允许我们进入他们的教室,并与他们的学生签约,这对我们很有帮助。第一步就是找到有大量学生选课的课程,并联系该课程的授课教师。我们发现教师在他们上课之初都非常配合我们的招募工作,于是我们特地早一点到。通常我们会先做一个简短的说明,解释实验,以及是否有报酬或者参与后能否获得学分,然后再分发表格。我们的联系表格上要求填写姓名、电话号码、电子邮件地址,以及之前提及的实验中要求填写的信息。

手头有了一系列潜在的被试,下一步任务就是安排日程。我们使用日程调度员来安排实验参与者:获得一堆的联系表格以后,日程调度员会打电话或发邮件通知潜在的被试,并在指定的时间安排所需数目的被试前往实验室。通常情况下,日程调度员负责向实验室给出清晰的指导,确保每个被试出现在正确的时间和正确的地点。由于事先安排好的学生很少会全部真的出现在实验室,因此需要多个被试的实验必须安排多于所需人数的被试。由于可能在研究开始前一周就已安排好日程,因此我们还要在研究前一晚打电话提醒潜在的被试。

在我们系,实验者一起建立了一个常用被试库,所有的被试都可以从中提取。重要的是,如果在被试参与的实验中没有使用欺骗,那么在另一个实验项目中还可以再次使用这些被试。由于招募是一个相当费时的任务,因此多次安排同一被试是非常可取的做法。

社会学和心理学研究者很少从一般人群中招募被试。米尔格拉姆(Milgram,1974)是一个例外,他通过报纸广告从耶鲁附近的社区中招募成年被试。将来,招募可能会发生根本性变化,因为公开招募被试的基于因特网的实验无疑会越来越普遍。在网上进行的这些研究,任何有兴趣的人都可以参与其中,因此有可能以最小的代价或无需代价就可以获得大量数据。只有未来才会告诉我们实验控制是否会持续,在网上的公开招募中被试的权利是否还会受到保护。

现在让我们探讨一下被试的酬金问题。有实验(由穆尔、布伦南和辛普森进行)较早地对获得酬金的被试进行了分析。医学研究者,尤其是生物伦理学的学生,写了许多给被试付酬金的优点和缺点。反对的一方认为,医学实验中一些对被试的明码标价是不道德的,因为酬劳使被试更有可能遭受风险,而在没有酬金时这种风险他们是可能避免的(McNeill,1997)。一些人认为酬劳被试违反了告知同意原则,因为获得酬金可能会使被试忽视他们可能受到的风险。如果潜在的被试很贫穷,那么这两个问题都会更严重(McNeill,1993)。那么,如果在被试完成了研究后,再付给他酬金呢?这样,酬金是合乎道德的吗?

站在赞成的一方,威尔金森和穆尔(Wilkinson and Moore,1997)对酬金进行了有力的维护,不过在此我们不详述他们的观点。相反,我们要指出的是,大多数(但可能不是全部)社会科学研究,支付给被试的酬金不足以让人忽视有关风险的信息或违反告知同意原则。而且,在大多数社会学研究中被试的风险很小,尤其在今天更是如此,因为各地的 IRB 对保护人类被试非常敏感。因此支付酬金给社会学被试是合乎道德的,但是拒绝向没有全部完成实验的被试支付酬金却是不道德的。

人为因素:需求特性和实验者偏差

设计实验旨在产生结果。经验主义者的设计运用米尔的求异法,如果成功的话,就能发现结果与一个或多个实验控制变量之间的规律。理论驱动型实验的目的在于构造由理论建模结构和过程的可重复实验(Willer,1987)。但是,一些结果却是实验中的人为因素(artifact)。克鲁格兰思奇(Kruglanski,1975:103)将**人为因素**描述为“关于所观察到结果的原因的推理错误”,更为准确的定义是,人为因素是系统性的结果,产生于①调查者未施加控制的变量;②受控的变量产生了研究者或理论预期之外的不同结果。

人为因素会导致研究者做出错误的推理,因为它们影响了实验结果。在某些情况下,人为因素的结果与测量理论结构的变量结果之间的差异微乎其微。需求特性(demand characteristics)(Orne,1962)和实验者偏差(experimenter bias)是人为因素的重要来源。下文中提到这两者时,我们会做出定义。我们现在来说明,在实验设计中研究者可以通过三种方式使需求特性和实验者偏差减到最少。首先,他们可以在设计阶段仔细关注这些问题;第二,他们可以在实验协议中建立检测人为因素来源的方法;第三,他们可以系统地对新设计进行前测。

需求特性

奥恩创造了术语"需求特性",来描述在实验情境中无意间泄露研究假设的所有暗示(Orne,1962:779,1969)。奥恩指出,实验者—被试的关系可以使被试预料到"好被试"应该有怎样的行为表现,即使他们并不确定好被试有哪些具体的行为。好被试关心实验的成功,关心研究假设能否**得到证实**,结果就产生了**被试偏差**(subject bias)。一个根据需求特性采取行动的被试会有意或无意地使研究结果出现偏差。

实际问题比奥恩讨论的暗示更复杂。至少有以下四种可能,可以列入需求特性:

1. 被试正确地推测实验假设,并试图产生预期的结果。这是奥恩所描述的情况。

2. 被试正确地推测实验假设,并试图产生相反的结果。韦伯和库克(Weber and Cook,1972)提出了这种可能。

3. 被试错误地推测了实验假设,并试图产生他们推测的结果。

4. 被试错误地推测实验假设,并试图产生相反的结果。

根据这四种可能,需求特性产生的结果可能与预测结果一致,也有可能不一致。无论是否一致,需求特性都会导致错误的解释。假设需求特性产生的结果与理论预测一致,则无论实验情境是否会产生结果,研究结果支持理论预测的这一结论都

是错误的。

产生不一致结果的需求特性也会导致错误的解释。试想这样的情境:变量在实验控制下产生与理论预测一致的结果,但是需求特性产生强度相等却不一致的结果,形成了抵消。这样的结果会使研究者得出结论,理论的变量无效,结果不能支持理论。另一种情况下,需求特性可能产生了强烈的结果,甚至盖过了实验控制的结果。这样,不知情的研究者会得出结论:理论变量有效,效果恰与理论的预测相反。

根据奥恩所述,需求特性不会消除,因为实验者不能控制被试带入实验的所有兴趣。威勒(Willer,1987:191)明确了产生人为因素的需求特性所需的三种必要条件,我们将之前讨论的四种可能也考虑在内,对之进行了修改:①被试必须推断一个假设;②被试必须想产生一个与假设一致(或不一致)的结果;③被试必须能够产生结果。奥恩关注第一个条件,认为通过事后情况说明进行检测是验证需求特性的最佳方法。检测是非常重要的,但是研究者也可以采取措施减少和管理需求特性。我们回到科恩等人(Cohen et al.,1996)的研究,说明他们的设计如何通过消除偏差产生的机会来控制被试的偏差。

与我们一直使用的人为因素的广义概念一致(例如:未受控变量的结果),科恩等人的设计包括一些消除人为因素潜在来源的特征。第一,研究者认为"离开家乡"上学的人与留在家乡上学的人相比,或许在很多方面有所不同。因此,实验排除了密歇根大学的本地学生。第二,犹太学生也被排除了,因为研究者认为犹太文化传统是影响南部和非南部人行为的另一个来源[10]。第三,侮辱方式始终保持一致,并且总是发生在实验助手与被试有了身体接触(碰撞)后。第四,只要**消除产生偏差的机会**,被试就不会意识到侮辱(实验控制下的变量)是实验的一部分。因为他们没感觉到,就没有会影响实验结果的针对实验的态度,需求特性就不会产生效果。

还有其他一些基于需求特性的设计,用于减少人为因素的

---

10 调查者也排除了黑人和拉丁美洲人,以消除种族和道德暗示的可能性。

可能性。当被试独自参与一项实验时,就不可能区别由独立变量(或初始条件)引起的行为和由需求特性引起的行为,后者的动机是想证明或证伪真实或想象中的假设。但是,应考虑到在任何社会学实验中,被试都会像在之前章节所描述的交换网络研究中所做的那样出现交互行为。假设那些实验组中的**所有**被试都推断过预测的结果——理论预测的交换率。进一步假设,所有担当"好被试"角色的人都通过**独自角色扮演**(role-playing alone)的方式产生结果。他们就会实施所预测的交换率,但是不会产生出复杂的交互过程来引导他们。网络交换研究中的交换率是通过实验者熟知的交互过程产生的。只有两种场景下可能出现进行角色扮演的被试。第一,交互过程是空缺的:被试没有经过协商就设置预测率;第二,交互过程会进行得仿佛被试们在进行一场拙劣的表演。

遵循网络交换理论所描述过程的交互,与通过参与角色扮演来确定假设的被试之间的交互相比,有很大不同。角色扮演型交互不能排练,并且被试们一般都不是有才华的演员。因此,其结果将不会像是协商。任何人如果怀疑角色扮演型交互是否有所不同,应该提前参加短剧排练(或旁听),并评价扮演角色的演员表演的真实性[11]。

实验者偏差

需求特性是引起被试偏差的真实或想象的实验情境特征。实验者也可能无意中影响被试的行为。通常情况下,研究者进入实验室时完全了解研究假设,而且在多数情况下,研究助理也是了解的。一个了解假设的实验者会在谈话、站立或走路时以某种方式透露有关假设的信息。当实验者的口头或非口头暗示(无论含蓄与否)影响到被试的行为时,就会对实验造成

---

11 我们描述的情况与奥恩(Orne,1962)关于实验前期调查的观点或非实验(Willer,1987)的观点有很大的不同。非实验要求被试库的成员像在实验中一样表现(例如:角色扮演)。此处描述的情况更像是即兴的戏剧,没有任何角色描述的帮助。我们在第1章和第4章的论述中指出,在简单三人网络中成功协调的可能性很小。威勒所描述的复杂、多角色结构中成功协调的可能性更小(Willer,1999)。

实验者偏差。像需求特性一样，实验者偏差会使解释研究发现变得困难或不可能。

罗伯特·罗森塔尔（Rosenthal，1966）用一个实验来说明实验者偏差的影响力。三组学生协助开展一项研究，研究对象是在迷宫中跑动的实验室老鼠。一组学生被告知他们的老鼠水平一般，第二组和第三组被分别告知他们的老鼠很聪明和很愚笨。实验愚笨老鼠的学生一致报告了很差的表现，而实验聪明老鼠的学生报告了出色的表现。在实验室外，这样的偏差也会引起很严重的实际后果。罗森塔尔和雅各布森（Rosenthal and Jacobson，1968）研究了教师期待，并发现教师如果认为学生表现差，他们就有差的表现。而教师如果认为学生表现好，他们就会有出色的表现。

米尔格拉姆的服从实验就是在实验者偏差中进行的。似乎米尔格拉姆没有提出假设就开始了实验（当然，他也没有理论），但他与被试之间的积极交互似乎与高度的服从相关[12]。只有在以下两种情况下顺从是可以忽略的：①普通人对被试发出指令；②被试可以选择电击的程度。

通过开展“盲实验”，可以完全消除实验者偏差的可能性。如果实验者不知道研究中的实验处理，那么该实验就是“盲实验”。例如，在上文描述的科恩等人（Cohen et al.，1996）的研究中，侮辱人的实验助手和进行观察的实验助手都不知道被试来自哪个地区。

盲实验的变式——双盲实验，确保实验者和被试都不了解研究情况。斯坦利·沙克特和杰罗姆·辛格在早期的情绪状态的心理要素实验中使用了双盲实验设计（Schachter and Singer，1962）。沙克特和辛格告诉被试他们正在研究维生素注射会如何影响视力。被试被征询他们是否同意注射“Suproxin”，95%的人同意。在注射后，他们和其他被试一起被带到等待室，并被告知要等待20分钟，直到维生素生效。

12 要提醒读者的是，只有执行电击发生器的最高级别电击的被试才被归类为服从的。

沙克特和辛格实际上研究的是,社会情境如何影响被试被唤醒的情绪状态。实际上,被试被注射的是肾上腺素或安慰剂,肾上腺素会产生一种生理唤醒状态。被试被分为三组:①不了解注射会如何影响他们;②得到注射会如何影响他们的准确信息;③得到注射会如何影响他们的错误信息。房间中的第二个“被试”是助手,要求他表现出以下行为:玩笑性的(欢欣状态)或者愤怒状态。被试没有意识到研究者正在研究他们在与助手交互后产生的情绪。助手不知道被试是否注射了肾上腺素或安慰剂,也不知道实验者对于注射的后果作过什么说明。双盲实验的设计从根本上减少了实验者(或被试)偏差。

另一个解决实验者偏差问题的方法是,分析被有意引入的偏差的效果。威勒有意在交换网络研究中引入实验者偏差,并在最终状态检测它们的结果——交换率,所研究的结构就像布伦南研究的强大结构一样。通过排除法,交换率会快速接近中心位置。实验者给了进行实验的助手一系列消极提示语句,会根据被试同意的交换率而随机运用。当交换率**接近**预测值时,助手就会低下头并开始咕哝“现在我遇到麻烦了,大麻烦”。然后他继续说“这是一个灾难”。接着,“那就是,我就要失业了”。如果数据**远离**预测值,实验助手就会说出积极提示。

实验是否对实验者偏差敏感很重要,如果敏感的话,敏感程度是多大,知道这些是很有用的。有意引入实验者偏差的好处是,可以获得在有偏差的情况下,实验的脆弱性的量化值。正如威勒的报告所言(Willer,1987:198ff.),之前讨论的设计几乎可以完全保证不受实验者偏差的影响。暗示会慢慢减缓力量过程,然而最终会接近预测终点。只有当被试不再根据交换的点数获得酬金,而是由于“愉悦实验者”而得到酬金时,力量过程才会停止,并在某种程度上逆转。

检测和减少人为因素

研究者可以通过着眼于潜在困难,精心设计实验来减少需求特性和实验者偏差。他们可以建立检测潜在人为因素源的程序,并将它们加入实验设计;他们还可以严格地前测实验设

计。实验研究者通常会进行前测，在研究开始前系统地尝试实施实验。通常的做法是，从被试库中挑选人员，然后通过充分的排练进行实验。前测可以在任何一个时间点停止，并对"被试"进行提问和事后情况说明。我们用一个案例来进行说明，它是我们进行的一系列研究中最早的一个设计。

沃克和泽尔底切(Walker and Zelditch,1993)进行了许多实验，研究个体对于结构性经济失衡的反应。在一系列研究中，被试要进行一个解谜任务:5 人一组的被试得到 10 个谜语，每人得到 10 次尝试机会。每位被试都会得到一些对解谜有用的信息，但是，如果不拿到团队中所有其他成员的信息，被试就不能解出谜语。每正确解出一个谜语(每道谜语最多有 5 个正确答案)，被试组就将获得报酬，报酬在 5 人中平分。

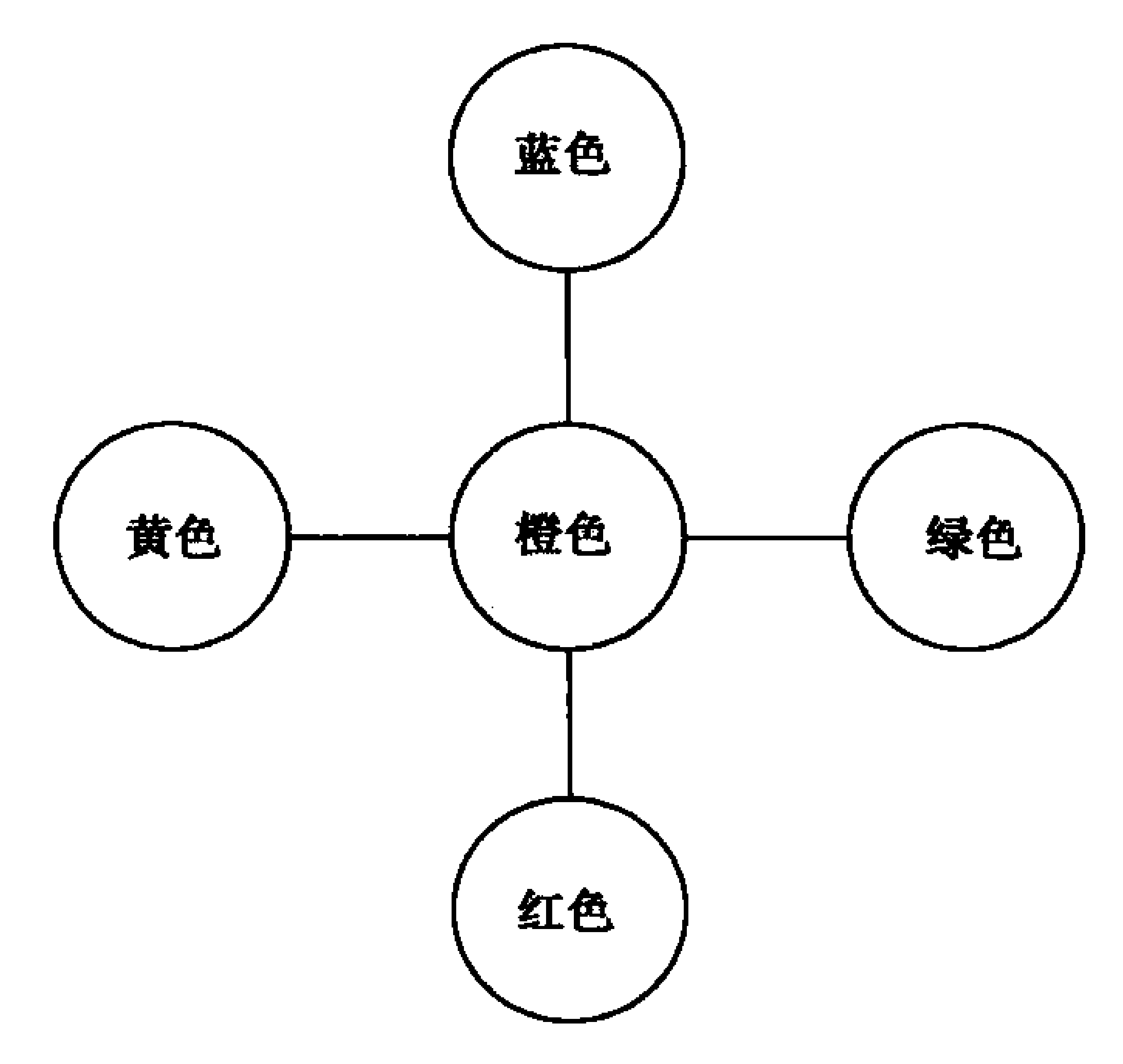


图 5.1 一个中心化的"轮子结构"

一些小组按照如图 5.1 所示的"轮子结构"开始工作，但小组成员被告知大部分人(3 人)可以改变结构的形状，并且还被告知如何进行改变。在最早的研究中，被试认为他们在通过相互传递手写消息来发送和收到谜语信息，但是消息实际上是由研究人员进行传递的。在解出了几个练习谜语并熟悉系统之后，实验者建立了一个激励机制促使成员更快地工作。10 个谜语中每一个首先提交正确答案的人，会得到额外的奖励。额外奖励形成了竞争，最初的合作任务变成了混合动机的、合

作一竞争的任务。虽然中心化的轮子结构对于这种交流而言是最有效的结构,但这种结构确保了处于中心的参与者(图5.1中的橙色)会获得所有的额外奖励。总体的不平衡对周围参与者形成了压力,使他们试图改变结构。一组研究探索了中心参与者对于试图改变结构的反应。

沃克和泽尔底切运用政治社会学和政治科学的理论提出假设:中心参与者会抵制结构的变化,因为这样的变化会削弱他们控制交互的权力,并且伤害他们作为理性参与者的利益,同样也会在实质上减少他们的所得。对处于中心地位的被试情境进行前测,结果表明几乎所有人都确实反对改变结构的意图。数据同时还显示,他们经常不能首先提交答案,因此许多次都"失去"了额外奖励。

事后采访发现,被试的动机与调查者的假设有很大不同。中心被试之所以反对变化是因为他们认识到结构不平等,并认为不公平,但是他们担心变化会使其他成员获得控制权,使奖励拱手送人。他们的解决方法是反对变化,保留权力,这样他们就可以确保每人都获得公平的奖励份额。当处于信息网络的中心时,他们可以决定其他的成员中哪一个先得到所有谜语信息,并等待那个成员提交他的信息,并将信息提供给其他成员,这样所有人都能得到正确的答案。他们的"轮换"办法可以确保他们和其他每个成员都能获得几乎相等的酬金份额。如果没有前测和事后采访,沃克和泽尔底切就不会知道被试的动机,也不会重新设计实验来去除被试对平等和公正的顾虑。

实验情境和以计算机为中介的实验

本节会为你介绍两个完全不同的以计算机为中介的社会实验研究情境。它们都是社会学中使用的"实验装置",就像用于物理学中的线性加速器和大型分光计。也就是说,它们产生实验条件,并测量结果。这些对于你来说都不会完全陌生,因为它们是你在之前章节中所见过的实验中用到的装置:验证地

位特征理论和基元理论的实验。对于这两个理论，检测的目的是为了增加实验者对被试所遇到情境的控制，以免超出先前设置。实现的目标：①理论模型和实验复本越来越相似；②减少实验者和被试偏差；③复本与初始情况相同。我们会向你解释如何在你的研究中使用这些装置。

以计算机为中介的实验的出现，从根本上改变了进行社会学研究的实验室。例如，我们几年前使用过的堪萨斯大学的实验室，原本是一个办公室，临时划出用于进行实验。房间内有椅子、成堆的扑克筹码——可以作为协商的筹码、一个煮蛋计时器——实验者可以用于计算协商回合时间，用夹板分隔成网络形状，从而阻止一些交互——那些未被阻止的交互形成了实验网络。实际上，作为布伦南实验讨论部分的一部分，在第4章中描述了那个实验室的布局。

南部卡罗莱纳大学的社会学研究实验室，是开展计算机为中介的实验的实验室布局的最好例子，我们在那里曾经进行过合作性实验。它更大，并且完全不同于堪萨斯的实验室。如图5.2所见，有12间小房间，包括2个控制站，围绕在大型中心房间的外围。每个小房间都配有一台PC机，而大房间每边配有八台以上的PC机。实验将被试安排在小房间，除了通过实验装置作为中介以外，被试之间无法进行联系。对于更大的网络，被试也会被安排在两个贴着“控制站”图标的小房间中，随着实验控制移到一个大房间的一台PC前。

我们现在来关注两个实验设备，在这一过程中，我们要求你首先记住验证理论的实验必须根据理论进行设计。事实上，现在将要描述的每台实验设备都是根据它要检验的理论进行设计的。你会注意到穆尔实验的设备和验证地位特征理论实验使用的设备之间的相似性，你也会发现布伦南实验设备和验证基元理论使用的设备之间的相似性。为什么会有这些相似呢？原因不在于设计电子仪器时复制了之前实验的设备，而是理论产生了这些相似性。每个理论早期的装置和电子设备相类似，因为它们都是根据理论建立的。

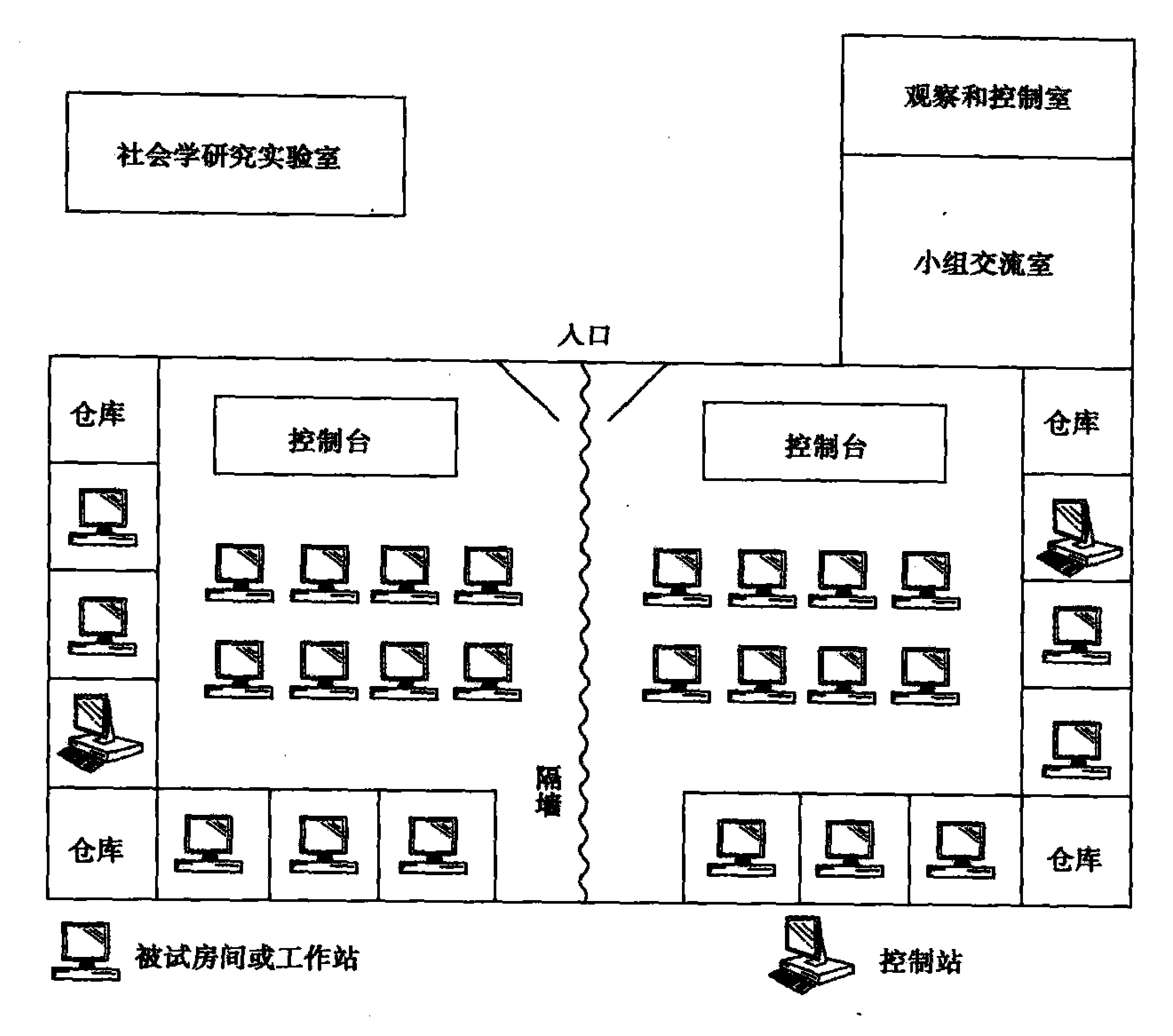


图 5.2 南部卡罗莱纳州大学实验室的布局

地位特征实验的电子设备

爱荷华大学的莉萨·特罗耶开发了我们现在描述的装置，使你和其他感兴趣的学者能够免费使用。但是，她要求那些有经费的项目捐款给爱荷华大学的进程组中心。捐款将用于支持学生研究项目。

使用这款软件，就有可能进行新的地位特征理论(SCT)实验或复制穆尔的(Moore,1968 见第 4 章)或许多其他的 SCT 实验。电子实验的初始条件设置与穆尔的做法非常相似。还记得 SCT 的验证需要这样的情境:参与者们的地位特征有一处或更多处不同，他们使用地位信息组织彼此的交互。被试坐在一台 PC 机前，它提供一整套任务指令。为了产生共同任务目标，装置会通知被试:①他会与一个同伴一起工作;②目标是针对

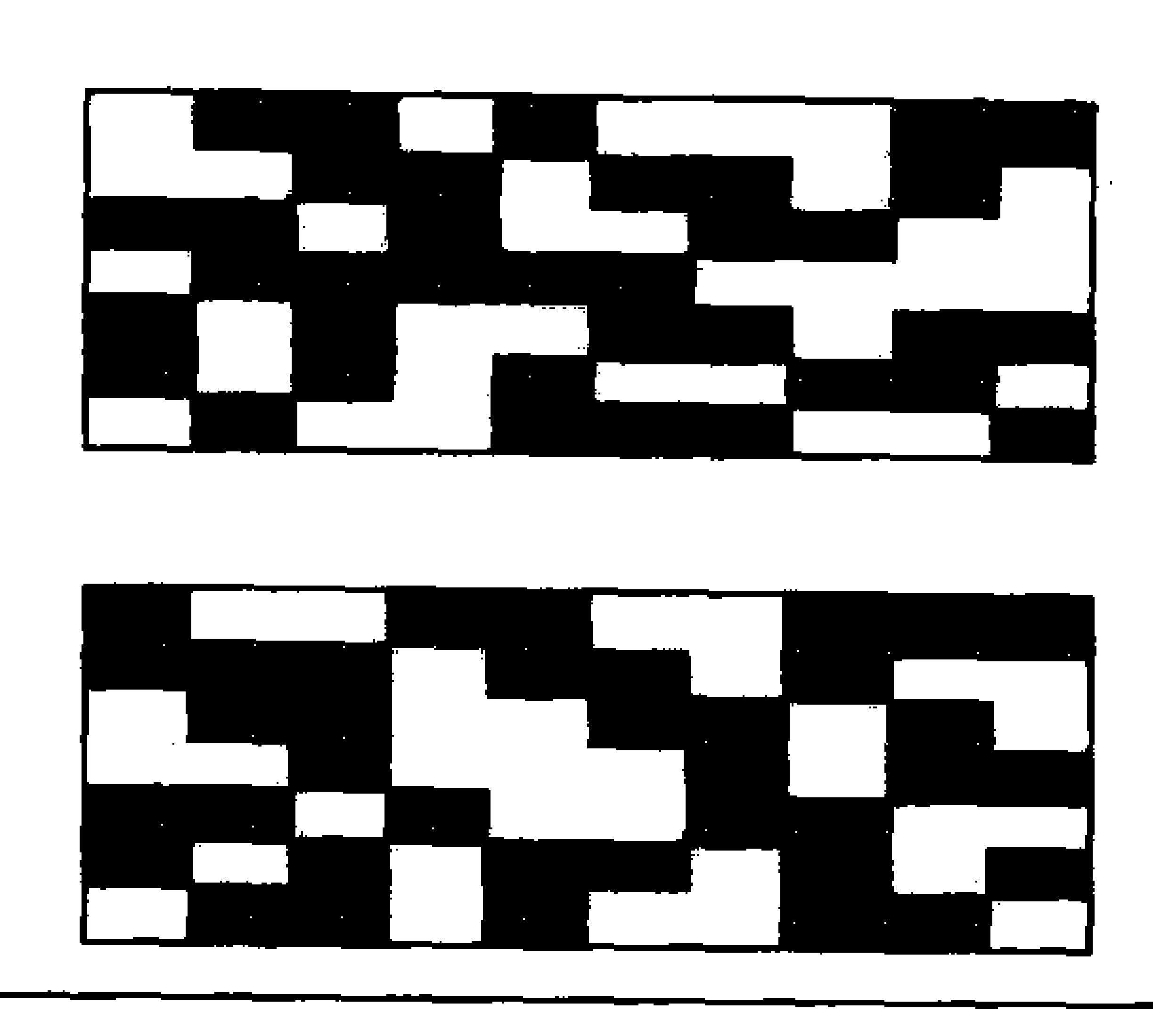
显示在电脑屏幕上的问题给出最佳回答;③结束时他会获得同伴的回答信息。

被试还会收到信息,使他们能比较自己和其他人的某些地位条件。例如,如果平均分为 C+ 的 21 岁大学生是低地位水平,她会被告知她的同伴是一个三十岁的研究生,平均分是 A。为了增加实验的真实性,被试会得到一张她的同伴的快照,同伴指的是用于代表高地位或低地位的个体。任务指令保证了被试不会在研究之后见到他们同伴。

像穆尔的研究一样,每个实验包含了一位真实的被试和一位虚拟的被试,但这里虚拟被试的角色不是由实验者扮演,而是由电脑扮演。装置会向被试展示对比敏感任务(Contrast Sensitivity Task)。被试在他的 PC 机上浏览了一系列如图 5.3 所示的幻灯片。正如所见,两个矩形,一个在另一个之上,有差不多相等的黑白不规则图案。在幻灯片下面是三个矩形的控制板,每个控制板的每条边上都各配有一个灯和一个按钮。第一个控制板上贴着一个标签,居中标有“你的初始选择”,左面的按钮标着“顶部”,右面的按钮标着“底部”。第二个控制板标有“你同伴的初始选择”,第三个控制板标有“你的最终选择”,后两个控制板的左右按钮也分别标着“顶部”和“底部”。

幻灯片出现 5 秒左右,然后,被试控制鼠标作出初次判断——顶部或者底部的矩形中是否有更多的白色(或黑色),并通过点击“初始选择”控制板两个按钮中的一个来表明自己的判断。按下按钮后,一个相应的灯会亮起,随后隔一段时间,PC 机会亮另一个灯,代表由虚构同伴作出的选择。在被试作出初始选择后的一段延迟时间是计算机中介呈现虚拟被试作出决定所需的不同时间。然后幻灯片再现,屏幕指示被试通过第三个也就是最底下的控制板记录最终选择。

PC 程序确保了被试和他的虚拟同伴意见不同的次数符合实验者的设定。在穆尔实验中,被试做了初始选择,而为与(虚拟)同伴的选择保持一致又做了不同的最终选择。这些改变了的最终选择的次数就是对影响的测量。在标准实验中,被试坚持初始选择的频率是 p(S),这是坚持或自我反应的概率。像



控制板

项部　　你的初始选择　　底部

项部　　你同伴的初始选择　　底部

项部　　你的最终选择　　底部

图 5.3　SCT 实验中的双幻灯片对比敏感任务

以前一样,高 p(S)值代表了更强烈地拒绝影响,而低 p(S)则相反。SCT 暗示高地位的参与者在与低地位的参与者交互时,有更高的 p(S)值,而认为后者的 p(S)值较低。实验装置使研究者能够广泛调查特征(例如,性别、性取向、教育程度)以及各种特质(例如,音乐才能、运动才能或完全编造的能力)的影响。

哪种物理布局是 SCT 设备所需的呢? 记住,每个实验只有一个真实的被试,因此所需的并不多。但是仍需进行物理安排,使被试认为那里的确还有另一个被试参与实验。一种实现方法是,布局时安排两位被试同时实验,例如前面提到的进行社会学研究的 USC 实验室。但这样的大型实验室可能过大了,因为使用所有周围的小型房间,就可以同时进行 12 项研究。

基元理论实验的电子设备

不同于 SCT 实验，基元理论（ET）实验的所有阶段都有多个被试，从 2 个到 10 个或更多。实验装置——ExNet 使被试能够以 PC 机为界面进行互相交流。ExNet 实验系统的网址是 http://weblab. ship. edu，它可以在线使用，通过因特网连接被试的电脑。每一项研究都由实验者创建，他可以通过自已和被试们使用的 PC 机上的浏览器访问、激活软件。

设计 ExNet 旨在验证 ET：它建立的复本与由 ET 建立的模型相对应（它也可以用于验证其他发生在结构中的社会关系理论，例如交换。）。ET 模型的参与者被视为理性的决策者。如果被试要符合理论中对参与者的设定，那么每个人都要完全了解他所交互的结构。为此，每位被试的屏幕都要显示网络，指示被试的位置，显示所有活动，包括所有被试在结构中完成的提议、反提议和交流。另一种替代性处理是，呈现给每位被试的信息采用多种方式加以限制。

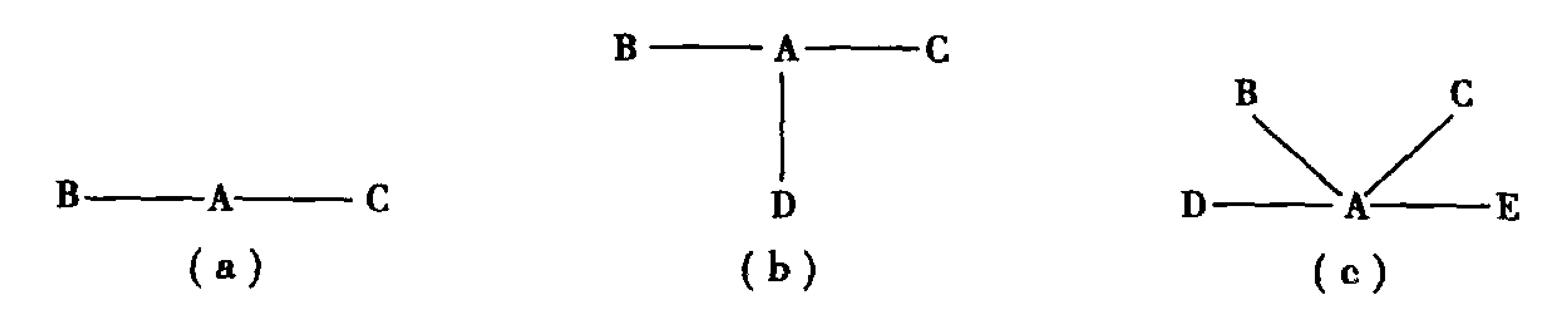


图 5.4 呈现给被试的三种网络交换结构

如图 5.4 所示，有三种结构。每位被试都能容易地找到他自己的位置，因为表示自己的字母比表示其他被试的字母大。图 5.4 中的三种结构都从被试 B 的角度呈现。当然，A 被试的屏幕上字母 A 最大，被试 C 的屏幕上字母 C 最大。如果被试是理性的，他们必须能掌控被试—PC 界面。因此，每个阶段之前都有指导性解释：如何提议、阅读他人的提议、完成交换。为了保证部分被试的专业知识，在实验之前，他们会在结构中练习，这不同于调查中的练习。

为了验证 ET 建模结构的所有范围，ExNet 必须产生一个结构化条件的序列。事实上，ExNet 可以调查所有结构化属性，并在最多容纳 25 个被试位置的任何结构配置中依据 ET 推进进一步发现。

每个 ExNet 阶段都会采取以下形式。当被试到来时,他们会被带到实验室外围的小房间中(见图 5.2),每一个小房间里都有一台已经联网并连接到 ExNet 的 PC 机。开始时呈现指导说明,随后是练习结构,然后是实验结构。在实验中,被试通过一系列回合进行协商,每一个回合都开始于更新过的资源。当所有能完成的交换都完成了,或没有完成但限定的时间已到,则该回合结束。实验者可以将回合细分为阶段,并在每个阶段的开始,用电子化方式轮换被试的位置。例如,在一个阶段里,每一位被试都会被安排在结构的各个位置轮流一次。

为每位被试提供完整的信息会使实验非常真实。经常可以看到被试虽然是一个人,但高度参与其中,大声说:"为什么我的提议没有被采纳!"以及"不,出的价钱太低了!"偶尔,非常激动的被试还会猛击墙壁。

一些设计特征被有意地引入 ExNet 中。首先也是最重要的,ExNet 的复本通常会完美地符合被考察的理论模型。例如,图 5.4 的结构是一个源于理论的网络配置模型,复本结构显示在被试的屏幕上。除了理论中所建构的和复本中所存在的,实验复本中不会产生交流。在图中,B 不能与 C 交流,C 也不能与 D 交流,同样的情况适用于所有不直接相连的被试位置。此外,交流的内容受到控制。每个被试可以自由地提议分配资源给与他相连的人,但不能发送威胁或"额外"收益。这些高度控制非常不同于布伦南的面对面实验中所建的模型和所发生的一系列交流。

ExNet 使许多其他重要的理论属性以非常纯粹的形式进行研究。被试的决策可以对应也可以不对应 ET 模型的理性决策。然而,在这里理性所需的条件包括完整的信息,训练有素的称职的被试,以及高度参与所需的持续关注。某个被试的特质诸如好的协商能力,可以通过轮换被试的结构位置而抵消掉。由于被试都会轮到有利和不利的位置,轮流安排位置会增加被试所得利益的平等性,从而降低了对平等性的担忧。

并非所有的大学都会贡献出自己的实验室,当他们没有这样做的时候,ExNet 实验会在计算机教室进行。只要计算机都连接互联网并安装了 IE 浏览器,那么计算机教室就可以用作

实验室。当然,计算机教师不能将被试分别隔离在单独的实验室中,但在有些情况下,这无关紧要。被试通过 ExNet 进行交互,忙碌地查看他们的屏幕,点击图标进行他们自己的提议、反提议和达到共识。他们只有很少的时间或几乎没有时间注意房间中的其他人。此外,被试可以相互看见或许没有一开始看起来那么重要。配对实验是例外,任何被试都不可能知道其他人在网络中的位置。

总 结

在各类学科中,反复证实了这样一个规律:将实验与理论相联系是使知识增长精确化、普遍化的最佳方式。最近,实验经济学知识的迅猛增长就表明了其(实验方法)能够达到的程度。更进一步来说,最近二十年,在社会学领域,基于实验的理论正在增长,不仅增长,而且逐渐积累。如果在社会学中实验研究像在经济学中一样增长,十年之内理论驱动型实验会成为我们领域的首要方法。

我们总结本书内容,归纳出三项重大问题。第一,我们考察了社会学中实验方法的应用,发现其应用比预想的更广泛——实验研究不仅限于小团体的研究。我们是终身的实验者,我们大多数的实验以解释大型社会结构为目的。第二,我们提出问题:当理论无法被实验验证时,哪种研究方法可用。我们解释了范围抽样的想法,说明理论驱动型历史对照研究如何采用类似理论驱动型实验的逻辑形式。第三,我们提出问题:调查和实验是否可以相联系。

6 受控制调查方式

本章考察了社会学中用于验证和应用理论的调查方式，我们称这些调查方式为“受控制调查”——向科学哲学家欧内斯特·内格尔(Nagel,1961)借用的一个术语。受控制调查中的“受控制”不是指受实验控制，而是指受理论控制。因此，理论驱动型实验是受控制调查，而有些其他种类的研究也是受控制调查，但它们很少或根本不受实验的控制。事实上，所有受控制调查逻辑上都类似于理论驱动型实验。因为即使调查者既不控制初始条件，也不控制观察结果的条件，所有受控制调查的结果也会相同。

在我们采取另一种受控制调查方式之前，我们要先回应关于社会科学实验的使用和意义非常有限的说法——“社会科学实验效用有限”这一观点由来已久。马克斯·韦伯在实验方面参考文献较少的情况下宣称，实验只能用于“少量非常特殊的案例”(Weber,[1918] 1968:10)。后来，伯吉斯(Burgess,1929)宣称社会学家与物理学家不同，不能将他们研究的现象带入到实验室——“社会科学研究的对象，像人、群体和机构等，只有在‘社区生活’这个实验室中才能被研究”(Burgess,1929:47)。更近一段时期，利伯森(Lieberson,1985;Lieberson and Lynn,2002)认为，实验在社会学中用处很小或根本没用。更糟糕的是，他认为将基于物理学实验而构造的方法论用于社会学是不恰当的。

这些断言都源于对科学，对理论在实验和其他研究中的地位的根本性误解。例如，伯吉斯错误地认为，其他领域中实验成

功的关键在于将实验室以外的目标带入到实验室的能力。正相反，成功的关键是理论，理论的应用在实验室与研究领域之间架起了一座桥梁。对于利伯森而言，“实验方法的关键特征是将被试随机分配到研究的条件中”（Lieberson，1985：14）。这种方法“源于经典物理学，该模式完全不适用于社会学”（Lieberson and Lynn，2002）。本书的读者们知道利伯森是被误导了。很难想象经典物理学可以随机分配什么，或者分配后将会得到什么。尽管费希尔在1935年引入随机分配，但在物理学中，300年前伽利略就已经很好地发展了理论驱动型实验。下面我们将揭示出韦伯的错误，即认为实验的使用是非常有限的。事实上，有些实验有助于解释理论中心的宏观结构。不过，韦伯没有固执己见，认为实验是有限制的，提出了“心理实验”，即一种基于实验逻辑结构的模拟。（Weber，[1918] 1968：10）。

本章分为三部分：首先我们会探索实验的效用，并发现它比批评者认为的效用大得多；其次我们会关注受控制调查的非实验方式，并解释它们的设计；最后提出问题——受控制调查的实验方式或非实验方式是否能一起使用，是否现在应该共同使用，进而推进社会学领域的发展，从而对本章和本书作一个总结。

理论驱动型实验的效用

出于以下3个原因，人们认为社会学实验效用有限：

1. 实验仅仅研究小结构，其时间跨度短。它们的结果如何适用于大而长久的结构呢？

2. 实验是简单的，它们如何适用于复杂的现象呢？

3. 实验缺乏外部效度。他们如何产生普适性的知识呢？

正如我们现在所表明的，这些都没有限制实验方法的效用。

规模是问题吗

规模问题经常成为“为什么实验对社会学的效用有限”的

原因。例如,内格尔提出了受控制调查的观点,因为社会科学家不能针对大规模且长期的过程实施实验,诸如工业资本主义的崛起(Nagel,1961:451-452)。但是实验真的无法说明工业资本主义的崛起吗?我们在此不会试图囊括整个过程,但会指出很多人(包括韦伯)所说的社会结构,都是其必要的组成部分,因为它影响到了法律制度。按照韦伯的观点,如果受制于独裁的、征用的法律制度,工业就不能发展。

包容资本主义的法律制度是能够建立起来的,如果封建时代晚期的城市公民被给予了这样做的权力——关键在于获得城市的自治权。那个时代的欧洲,市民通过挑起国王与当地领主的争斗以及二者与教会力量的争斗,而获得自治。城市自治包括为市民减税,市民有权起草和执行他们自己的法律。换种说法是,由于封建主义的分权,国王、领主和教会的统治权部分重叠,所以城市就将统治他们的权力拍卖给这三个强权中出价最高(即给市民待遇最好)的一方(即市民获得自治)。

随后,封建主义衰落,权力集中到了国家体系,但是统治权力间的竞争并没有结束。拥有流动资本的公司通过在不同国家间转移或威胁转移投资,挑起国家间的互相对抗,这种对抗一直持续到今天。今天,对抗仍在继续,但不只是国与国之间,也存在于美国的各州之间。例如,南卡罗莱纳州政府部门资金充足,他们的目的就是发展并出高价将公司引入到他们那里。因此,该政府部门经常比其他州或国家出价更高[1]。

图6.1显示了"强制"中心结构。在结构中,D必须从三个C中挑选一个作为他的"统治者",但是在选择前可以与三个C进行协商。C在与D达成一致前不会传达消极反应,因此不需要明确进一步的参数,就能作出三个重要的预测:第一,三个C将竞争成为D唯一的"统治者";第二,所有C都不会传达他们的消极反应;第三,由于竞争,统治剥削率将达到系统可能的最低限。其中,统治剥削率低意味着低税收和自治的机会。事实上,三个中的两个"统治者"会被排除掉,来自布伦南研究的单

1 丰田汽车公司在2005年夏初宣布它将在加拿大的安大略(Austen,2005)建立第二个装配厂。加拿大和安大略政府出价要高于美国的几个州。

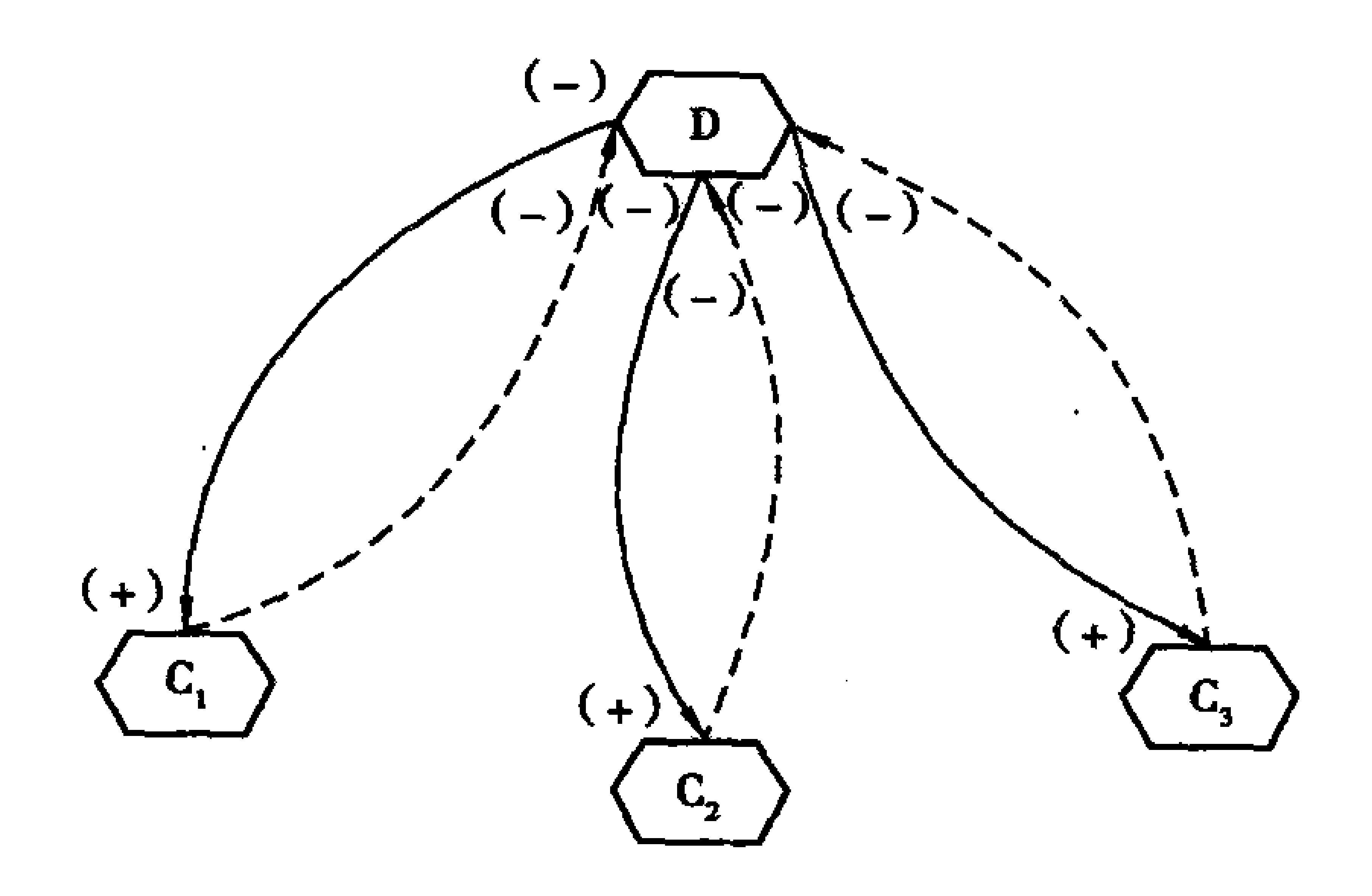


图6.1 “强制”中心交换结构

一联结交换结构的经验使得这三个预测毫无疑问地得到了支持。城邦与封建权力的斗争，流动资本家的公司与国家的斗争，这些是可以先在理论上建模的两个统治中心系统的案例。第4章论述的5个步骤，将实验和这些历史结构的其余部分连接起来，对“强制”中心模型做了复制，并拓展了研究。

有什么证据可以表明统治权力间的竞争是资本主义出现的必要条件呢？根据韦伯所述，资本主义不会在古代中国发展出来，尽管其技术比欧洲超前，而且中国人有强大的获取动机，但是在传统的中国，商业法不发达，国家征用发展中资本家公司的资产是很普遍的情况。为什么中国的商业法与资本主义的产生不相容呢？因为中国是一个中央集权的帝国。因此，与欧洲不同，古代中国不存在相互竞争的权力彼此抗争。这样就没有城市自治，也没有相应的法律体系（Weber，[1911] 1951）。这里还有一个相关模型，即不排他的统治，实验结果可以与以上描述的“强制”中心结构相对比。

或许第二个案例需要切中主题，规模不一定是障碍。如果内格尔关于对资本主义的发展不能用实验进行研究的阐述是正确的，那么同样——实验无法说明古代和近代美国大的奴隶结构兴起和衰落。与内格尔相反，统治结构的实验已经实施，而且它们支持了马克思和韦伯对于奴隶结构的看法。马克思（Marx，[1876] 1976：377）和韦伯（Weber，[1896] 1976：398）将统治剥削率，即（马克思认为的）利润，与廉价奴隶的供应相关联。在19世纪，在牙买加和巴西进口奴隶不受限制，与在美

国进口奴隶会触犯法律相比,能获利更多。他们的洞见被历史证明是正确的,但没有人给出原因——为什么利润与奴隶进口相关联。

基元理论预测统治剥削率应该会随奴隶主惩处奴隶的成本而发生变化。注意,惩处奴隶的成本与廉价奴隶的供应相关。当惩处奴隶的成本降低时(即廉价奴隶更易获得时),在某一点上奴隶就会竞争(即让自己比别的奴隶被剥削得更多)避免被惩处。竞争有利于统治者,就像在布伦南(Brennan,1981)研究的交换结构中:竞争对中心位置有利。如果统治剥削没有最高限额,竞争结果会非常惊人(奴隶会被剥削得非常惨)。验证该推导的实验为这一观点提供了强有力的支持(Willer,1987)。

很难想象同一结构会在规模上如此不同——一个是庞大的古代奴隶种植园或18、19世纪美国大型的奴隶庄园,一个是验证基元理论推导的小样本的实验性统治结构。这种(规模)差异对于无法理解理论驱动型研究逻辑的人来说就是问题。但理论的实验验证不需要科学家——无论是社会学家还是天体物理学家——在他们的实验室中重验模型规模的大小,所需要的是另一件事情——实验室和现场结构类似。如果我们从**理论视角来看**,统治中心化结构就像我们刚刚所描述的,看上去非常类似于奴隶结构。

除了规模以外,刚讨论过的历史事件和实验重验之间还有第二个重要的不同点:历史事件复杂得多。简单实验的结果和历史复杂性之间联系的理由是什么?

理论简单性和复杂的结构

理论可以将简单的实验复制与复杂的历史及当代社会结构相联系,因为它具有本书尚未讨论过的重要能力:**理论能创造出复杂的模型**。本书用于建立实验的简单模型可以进行结合重组,进一步创造出复杂的理论模型。由较简单形式组成的较复杂形式的整体,已经可以在社会关系分解成较小规模的结构中看到。以此类推,多种结构模型可以作为单个复杂模型的组成

部分。复杂模型可用于现场的解释。重要的是,它还可以用于建立大型的复杂的实验复本或模拟。遗憾的是,组合的方法虽然曾经被人提到过(Willer,1984),但今天在任何社会科学中几乎无人知晓。它的缺失是很奇怪的,因为在其他科学中这是无人不知(Einstein,[1934] 1954:27)的。此外,在社会科学中,组合的方法最早是由马克思提出的(Marx,[1857] 1973:10)。

组合的方法可行吗?社会理论是否足够先进以至可以联系较简单的模型,并且得出新结论?这里有一个假设的案例。如图6.2所示是两个相联系的交换结构α、β,C与α、β相连,即"资本家"占据了两者的中心位置(C就像一个拥有流动资金的现代资本家)。也就是说,C拥有所有的交换关系和两个网络中的位置,可以正向或反向地从α到β两个结构间的关系上自由移动。α的结构间无连接,如果没有β,交换是等量的。在β缺失的情况下,其工资会很高,利润会很低。由于无连接,α本身不会是一个有利于资本家的权力结构。

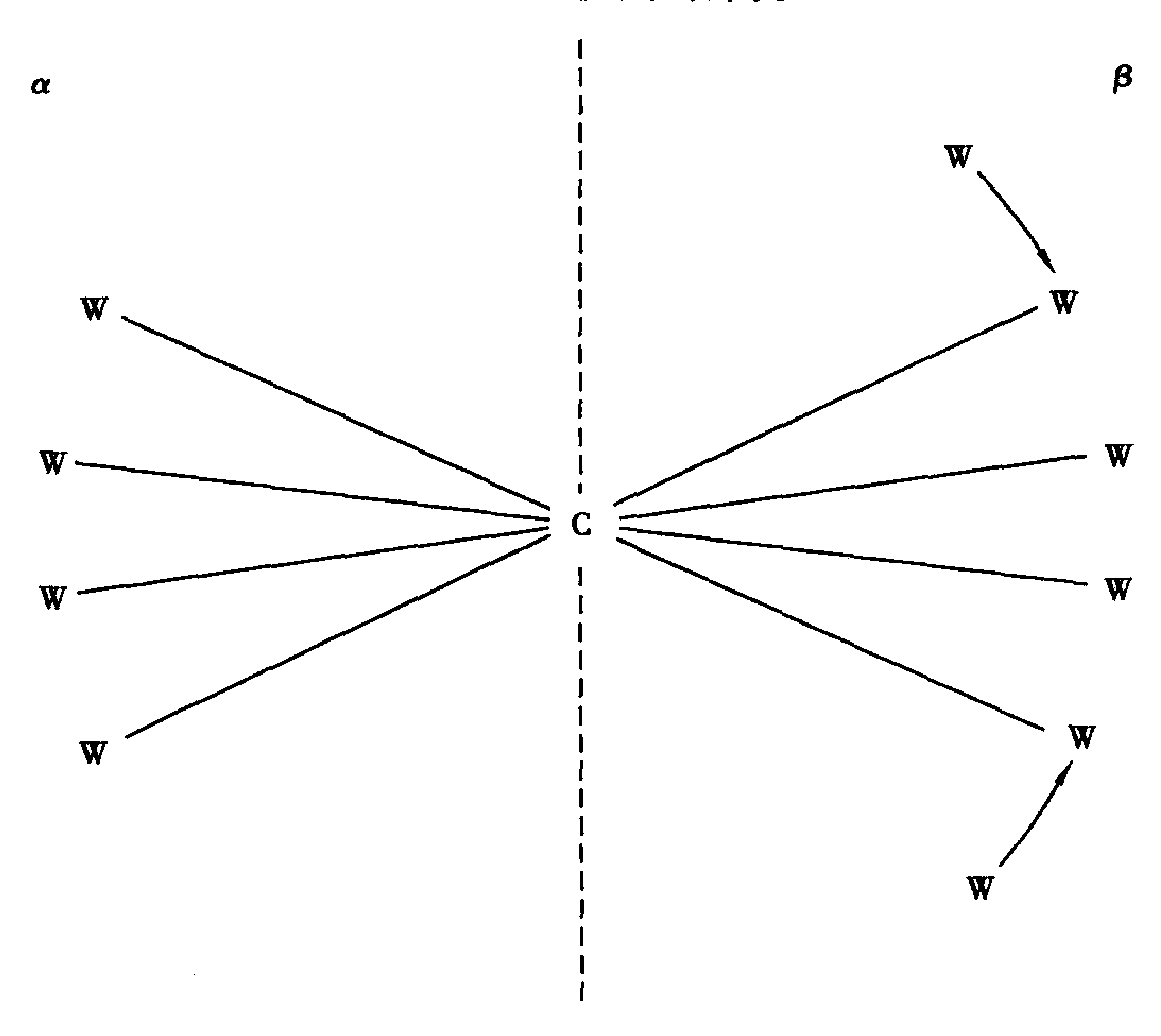


图6.2 两个相联系的交换结构模型

现在来看看β结构,它是一个因有多余的W参与竞争而使C位置拥有强权力的结构。多余的W无论数量多少都来自

α。由于β是一个强权力的交换结构,它的交换率极其有利于C。也就是说,利润是最大的,工资是最低的。由此立即出现了以下的新推论:C将提出确实可信的威胁——把交换关系从α移到β,而一个理性的、利己的C一定会这么做,除非α中的交换率像β中的交换率一样对C有利。由于模型中只有α和β结构,α中的W没有可选的工作,因此只能接受最低工资。

上文描述的模型提出,在复杂的网络中无连接的网络和有限连接的网络相连,因此无连接的网络会成为权力结构。一旦模型正式发表,研究者就会建立实验复本验证模型。此外,实验室外也有据可循。今天的全球经济体系中,资本高度流动,工作从美国转向了第三世界的国家,这威胁到了美国的工资。

建立和应用日趋复杂的模型的研究在今天是否可行?我们建议如此,正是由于理论研究中更难的工作已经在进行中。刚才讨论的复杂模型是单个理论(基元理论)的直接拓展。更难的研究是把两个理论联系起来。现在至少两项连接研究项目正在进行中。地位特征理论和基元理论由赛伊、威勒和马可夫斯基以两种方式联系起来,将影响过程和地位价值的扩散与交换率相联系(Thye, Willer and Markovsky,2006)。我们正在进行的实验研究将合法性理论(Legitimacy Theory)和基元理论联系起来,展示合法性水平影响到权力能否得到补偿。

理论、实验和外部效度

一些人宣称实验没有用,因为它没有外部效度。什么是"外部效度"?为什么有些人认为它很重要,并且重要得足以推翻实验研究?

"外部效度"意味着从随机抽取的样本中推断出总体参数的能力(见Lucas,2003的延伸讨论)。的确,推断总体参数时实验用处不大或没有用。但是这样的推断不是人们想要的,也不是需要的。理论驱动型实验的目的是验证一般而普遍的理论。我们所说的"普遍"是指不考虑时间和地点的具体条件下的理论应用。如果理论有解释和预测作用,必须是普遍的。相反,具有外部效度的样本仅允许推断出具体时间和地点条件下的总体。因为推断是具体的,而不是一般的现象,所以不能从

中起到解释或预测作用。

理论的规律不是对总体的估计。社会学理论的范围,像所有理论一样,都不能由可列举的集合加以定义,比如人口总体。它不像是调查中的抽样集合,任何理论范围中包含的数量都数不清,甚至不能罗列出来。相反,理论的范围是指现象的总和。此外,范围并不固定。当理论被验证时,有证据表明理论的范围比预想的狭窄。另外,随着理论的成熟和新形式的引入,理论的范围也扩大了。

自然而然,波普尔是正确的(Popper,1994)。与从样本向总体推断不同,理论验证不是也不可能是演绎。对于演绎而言,研究依赖于一系列实例(通常是人),依据随机样本推出结论。与之相反,理论应用的总体不可数,也不能被罗列出来。因此从中随机抽样是不可能的——因为它是归纳式的。与归纳不同,实验中推断的方向正好相反。在理论验证中,我们不是从实验归纳出理论,但可以从理论推导出实验。

受控制调查的非实验方式

还有两个问题。第一个问题关系到用于验证理论的实验室外的真实案例。在实验中,验证理论的经验案例是实验室创建的复本。在非实验的受控制调查中,必须找到验证理论的案例,更具体地说,必须选择它们。怎么控制选择呢?我们给出了"范围抽样"作为答案。第二,如何设计和执行每一项调查呢?我们给出的答案与问题"应该如何设计实验?"的答案是相同的。应由理论来设计并指导非实验的研究。

为验证理论而系统选择经验案例称为"范围抽样"。这个程序通过类比理论驱动型实验就可以很容易地理解。在理论驱动型实验中,我们构建了可在实验室被复制的实证案例。这些实验复本根据理论的初始条件进行排列。例如,地位特征理论(SCT)的实验,一般存在调查地位高的人(实验中他们认为自己是在与地位低的人互动)和地位低的人(实验中他们认为自己是在与地位高的人互动)。这些实验的初始条件就是根据

地位从高到低的范围进行排列的。与之相似,基元理论(ET)的初始条件是根据结构条件的变化范围、结构配置的变化范围或这两者一起排列复本[2]。布伦南实验改变了结构的条件,从无连接到有“结构之外的连接”;第1章的交换实验改变了配置,从三角形网络到中心化网络。

根据理论初始条件的范围排列复本基于两个同样重要的理由。第一,排列提供了非常出色的验证理论的能力,将初始条件和最终条件联系起来。在SCT中,我们想要表明,地位差异产生影响——地位排列复本的实验结果能够支持这一命题。在ET中,我们想要表明,权力随着结构条件和配置而变化——根据这两种条件排列的复本实验结果也能支持这一命题。第二,依据范围排列初始条件是对理论的最好验证,因为它标出了明确的理论范围,在这一范围中可以有效地预测和解释。此外,这是最严格的验证,因为它最大化了犯错和证伪的机会。

非实验研究中的范围抽样遵循与实验中的范围变化完全相同的逻辑。当对范围进行抽样后,研究者根据理论模型的初始条件中的变量选择实例。这里有一个范例,如前所述,统治中心化结构的模型可以采用历史事件加以验证。在古代中国这样的统一帝国中,统治关系到处可见,但是不存在统治中心结构。统治权没有重叠,也不存在小国之间的资本流动。在中国可以找到很高的统治剥削率。相比之下,欧洲有统治中心结构,首先是由于重叠的统治权,其次是由于欧洲各小国间资金流动的方便性。在欧洲可以找到低的统治剥削率。统治剥削率水平可以从两个方面进行测量:一是税率,二是法律体系对商业发展的适应性。

考虑一下第二个案例,将SCT和ET相结合。我们再次根据理论范围排列案例,这个案例是根据市场中买家的地位进行排列的。地位是否像SCT和ET一起预测的那样,反过来与支

---

2 如果理论有自变量和因变量,那么我们可以说引入自变量上的变异,以至于能追踪到因变量的变化。范围抽样的观点由来已久。一个“范围抽样”可以被定义为一系列分布在某一公认的体系中主要维度上的,符合理论模型的相应条件的真实案例(译者注,就像在彩虹中抽样出七种颜色,就是在从红色到紫色范围的抽样。)(Willer,1967:114)。虽然这一观点提出已有40年,但是我们了解到很少有研究使用这一程序。

付给商品的价格相关？例如，购买同一辆车时，黑人女性是否比白人男性支付更多？最近的调查表明，黑人女性支付得更多（Ayers，2006）。因而 SCT 和 ET 的联系得到了支持。

传统的同时代对比研究和历史对比研究的目的在于使用米尔方法的经验规则（诸如求异法、求同法和共变法），得出经验规律。由于米尔的方法没有提供区分哪些事物需要观察而哪些不需要观察的标准，而传统的比较研究是开放式的，因此无法理性地得出结论。例如，研究无法通过成功地发现规律而得出结论，因为经验表明方法不能揭示米尔所期望的绝对规律。社会学中最好的比较研究，像皮歇尔（Prechel，2000）和曼（Mann，1986）的研究，没有身陷于米尔的方法。他们正确地克服了米尔方法的缺点，因为他们超越了米尔，遵循了像马克思和韦伯这样的理论家的设计指导。这些设计指导着研究者什么应该观察，什么不该观察，什么关系可能被找到，什么关系不可能被找到。基于这种做法，他们使得比较研究合理地被执行，合理地得出结论。

尽管如此，运用于理论驱动型实验中的更为规范的理论的非实验应用，改善了由经典理论指导的研究，并产生了一种新的同时代对比研究和历史对比研究。同时代对比研究中，正式理论指导研究者该观察什么，不该观察什么。历史对比研究中，正式理论指导研究者应该挖掘出哪些数据来应用理论。在经验案例和理论间反复研究，第一个模型便随之建立，并由涵盖案例所需的日趋复杂的设计所构成。事实上，模型的范例应用在许多历史事件中，在前一部分中已经给出了大概形式，包括由两个相连接的结构组成的模型。

正如实验验证力求覆盖广阔的范围一样，范围抽样对理论作了严格验证，因为它使犯错和证伪的机会最大化。并且像实验一样，范围抽样广泛标出了明确的理论范围，在这个范围内可以有效预测和解释。重要的是，理论驱动型实验的逻辑和范围抽样的逻辑是一样的，因此，它们对于理论的验证同样重要。

实验和非实验调查的联系

实验受控制调查和非实验受控制调查方法可以结合起来提供比先前更精确、范围更广的预测和解释吗？本章已经表明了两种调查都是由同一个逻辑驱动的，甚至是由同一个理论驱动的。不仅结合两种类型研究的目标是可行的，而且朝向这个目标的第一步已经开始了——尽管案例还很少。除了第4章讨论的科恩的研究外，还有威勒及其同事(Willer et al.,1997)应用实验验证的模型来解释韦伯关于西方古代文明的衰落的观点；那个中心模型的应用，包括本章前面关于奴隶的两个强制结构；贝尔、沃克和威勒(Bell, Walker and Willer,2000)应用来自基元理论、地位特征理论以及合法性理论的模型到正式组织结构中，来解释在那里发现的高度顺从。

对于物理世界的理解在近百年来突飞猛进，因为物理学和天文学结合了实验和非实验研究，对于生物世界的理解在近五十年来突飞猛进，因为遗传学和进化生物学结合了实验和非实验研究。在将来的五十年，对人类社会的理解也将突飞猛进。而这将从人们习惯于社会学理论研究中实验和非实验研究方法相结合开始。

专有名称汉英对照表

人名汉英对照表

亚历克斯·贝弗列斯 (Alex Bavelas)
阿基米德 (Archimedes)
亚里士多德 (Aristotle)
阿施 (Asch)
培根 (Bacon)
贝尔斯 (Bales)
比利·巴德 (Billy Budd)
伯克 (Berk)
伯格 (Berger)
巴夫 (Bargh)
伯罗 (Burrows)
布伦特·辛普森 (Brent Simpson)
贝尔 (Bell)
布伦南 (Brennan)
伯吉斯 (Burgess)
鲍德尔 (Bowdle)
贝尔斯 (Bales)
陈 (Chen)
科恩 (Cohen)
坎贝尔 (Campbell)
库克 (Cook)
卡罗莱纳州 (Carolina)
科拉 (Corra)
笛卡尔 (Descartes)
达夫·科恩 (Dov Cohen)
笛卡尔 (Descarte)
达尔文 (Darwin)
欧内斯特·内格尔 (Ernest Nagel)
伊丽莎白·G. 科恩 (Elizabeth G. Cohen)
欧几里德 (Euclid)
费希尔 (Fisher)
弗里兹 (Freese)
费尔马 (Fermat)
盖伦 (Galen)
伽利略 (Galileo)
格瑞夫 (Graif)
伽利略 (Galileo)
亨普尔 (Hempel)

希尔罗 (Hieron)
惠更斯 (Huygens)
休斯 (Hughes)
爱荷华 (Iowa)大学
约瑟夫·伯杰 (Joseph Berger)
J. S. 米尔 (J. S. Mill)
詹姆斯·穆尔 (James Moore)
詹姆斯·C. 穆尔 (James C. Moore)
雅各布森 (Jacobson)
杰罗姆·辛格 (Jerome Singer)
库瓦巴瑞 (Kuwabara)
堪萨斯 (Kansas)大学
克鲁格兰思奇 (Kruglanski)
开普勒 (Kepler)
利伯森 (Lieberson)
莉萨·特罗耶 (Lisa Troyer)
劳德·汉弗莱斯 (Laud Humphreys)
莱迪亚德 (Ledyard)
勒尼汉 (Lenihan)
利希特 (Lichter)
卢卡斯 (Lucas)
洛瓦哥列 (Lovaglia)
麦克劳克林 (McLaughlin)
米尔格拉姆 (Milgram)
米尔 (Mill)
斯坦利·米尔格拉姆 (Milgram)
曼特尔 (Mantell)
马克思 (Marx)
马可夫斯基 (Markovsky)
马丁·奥恩 (Martin T. Orne)
穆尔 (Moore)
马克斯·韦伯 (Max Weber)
尼斯比特 (Nisbett)
牛顿 (Newton)
奥恩 (Orne)
波普 (Popper)
罗西 (Rossi)
皮列文 (Piliavin)
菲利普·津巴多 (Philip Zimbardo)
柏拉图 (Plato)
托勒密 (Ptolemy)
罗丹 (Rodan)
罗伯特·罗森塔尔 (Robert Rosenthal)
瑞巴 (Ribar)
R. A. 费希尔 (R. A. Fisher)
斯坦利 (Stanley)
塞尔利茨 (Selltiz)
塞尔 (Sell)
所罗门·阿施 (Solomon Asch)
西默尔 (Simmel)
斯坦利·沙克特 (Stanley Schachter)
施瓦茨 (Schwarz)
西伯 (Sieber)
斯内尔 (Snell)
辛普森 (Simpson)
赛伊 (Thye)
沃克 (Walker)
威勒 (Willer)
韦伯斯特 (Webster)

威廉·戈塞特 (William S. Gossett)
威廉·詹姆斯 (William James)
韦伯 (Weber)
沃克 (Walker)
威勒 (Will)
威尔金森 (Wilkinson)
泽尔底切 (Zelditch)

专有名词汉英对照表

《美国社会学协会道德守则》 American Sociological Association's Code of Ethics

机构审查委员会 American Sociological Association, ASA

《逻辑体系》 A System of Logic

贝尔蒙特报告 Belmont Report

过渡期援助研究计划 Transitional Aid Pesearch Project, TARP

《数学原理》 Mathematica

地名：

锡拉丘兹 Syracuse

亚历山大 Alexandria

参考文献

American. Anthropological Association. 1998. *Code of Ethicc of the American Anthropological Association*. Washington, DC: American Anthropological Association.

American Psychological Association. 2002. *Ethical Principles of Psychologists and Code of Conduct*. Washington, DC: American Psychological Association.

American Sociological Association. 1999. *Code of Ethics and Policies and Procedures.* Washington, DC: American Sociological Association.

Archimedes. [230 BC] 1897. *The Works of Archimedes*. T. L. Heath (ed.). New York: Dover.

Aristotle. [347 BC] 1962. *Nicomachean Ethics*. Martin Ostwald (tr.). Indianapolis, IN: Bobbs-Merrill.

——. [330 BC] 1961. *Physics*. Richard Hope (tr.). Lincoln: University of Nebraska Press.

Asch, Solomon E. 1958. "Interpersonal Influence." Pp. 174-183 in Eleanor Maccoby, Theodore Newcomb, and Eugene Hartley (eds.), *Readings in Social Psychology*, 3rd Edition. New York: Holt, Rinehart and Winston.

Austen, Ian. 2005. "Toyota Is Said to Be Planning Its Second Factory in Canada." *New York Times*, Late Edition, June 24, 2005.

Axelrod, Robert. 1984. *The Evolution of Cooperation*. New York: Basic Books.

Ayres, Ian. 2006. "Discrimination in Consummated Car Purchases." Chapter 6 in Laura Beth Nielsen and Robert L. Nelson (eds.), *Handbook of Employment Discrimination Research*. New York: Springer.

Bacon, Francis. [1620] 2000. *Novum Organon*. Lisa Jardine and Michael Silverthorne (eds.). Cambridge, UK: Cambridge University Press.

Bales, Robert F. 1950. *Interaction Process Analysis*. Cambridge, MA: Addison-Wesley.

——. 1999. *Social Interaction Systems*. New Brunswick, NJ: Transaction.

Bargh, John A., Mark Chen, and Lara Burrows. 1996. "Automaticity of Social Behavior: Direct Effects of Trait Construct and Stereotype Activation on Action." *Journal of Personality and Social Psychology* 71:230-244.

Bavelas, Alex. 1950. "Communication Patterns in Task-Oriented Groups." *Journal of the Acoustical Society of America* 22:725-730.

Bell, Richard, Henry A. Walker, and David Willer. 2000. "Power, Influence and Legitimacy in Organizations: Implications of Three Theoretical Research Programs." Pp. 131-178 in Samuel B. Bacharach and Edward Lawler (eds.), *Organizational Politics*. Stamford, CT: JAI Press.

Berger, Joseph, Bernard P. Cohen, and Morris Zelditch, Jr. 1966. "Status Characteristics and Expectation States" Pp. 29-46 in Joseph Berger, Morris Zelditch, Jr., and Bo Anderson (eds), *Sociological Theories in Progress*, vol. 1. Boston: Houghton-Mifflin.

——. 1972. "Status Characteristics and Social Interaction." *American Sociological Review* 37: 241-255.

Berger, Joseph and M. Hamit Fisek. 1970. "Consistent and Inconsistent Status Characteristics and the Determination of Power and Prestige Orders." *Sociometry* 33:287-304.

——. 2006. "Diffuse Status Characteristics and the Spread of Status Value: A Formal Theory." *American Journal of Sociology* 111:1038-1079.

Berger, Joseph, M. Hamit Fisek, and Paul Crosbie. 1970. "Multi-characteristic Status Situations and the Determination of Power and Prestige Orders." Technical Report #35. Stanford, CA: Laboratory for Social Research.

Berger, Joseph, M. Hamit Fisek, and Robert Z. Norman. 1998. "The Evolution of Status Expectations: A Theoretical Extension." Pp. 175-205. in Jospeh Berger and Morris Zelditch, Jr. (eds.), *Status, Power and Legitimacy*. New Brunswick, NJ: Transaction.

Berger, Joseph, M. Hamit Fisek, Robert Z. Norman, and Morris Zelditch, Jr. 1977. *Status Characteristics and Social Interaction*. New York: Elsevier.

Berger, Joseph, David Willer, and Morris Zelditch, Jr. 2005. "Theory Programs and Theoretical Problems." *Sociological Theory* 23:127-155.

Berger, Joseph, Morris Zelditch, Jr., and Bo Anderson. 1972. "Historical and Generalizing Orientations in Sociology." Pp. *ix-xxi* in Joseph Berger, Morris Zelditch, Jr., and Bo Anderson (eds.), *Sociological Theories in Progress*, vol. 2. Boston: Houghton-Mifflin.

Berk, Richard A., Kenneth J. Lenihan, and Peter H. Rossi. 1980. "Grime and Poverty: Some Experimental Evidence from Ex-Offenders." *American Sociological Review* 45:766-786.

Brennan, James. 1981. "Some Experimental Structures." Pp. 198-206 in David Willer and Bo Anderson (eds.), *Networks, Exchange and Coercion*. New York: Elsevier.

Burgess, Ernest. 1929. "Basic Social Data." Pp. 47-66 in T. V. Smith and L. D. White. *Chicago: An Experiment in Social Science Research*. Chicago: University of Chicago Press.

Campbell, Donald T. and Julian C. Stanley. 1966. *Experimental and Quasi-Experimental Designs for Research*. Chicago: Rand McNally.

Cohen, Bernard P. 1989. *Developing Sociological Knowledge*, 2nd Edition. Chicago: Nelson-Hall.

Cohen, Dov, Richard E. Nisbett, Brian F. Bowdle, and Norbert Schwarz. 1996. "Insult, Aggression and the Southern Culture of Honor: An 'Experimental Ethnography.'" *Journal of Personality and Social Psychology* 70:945-960.

Cohen, Elizabeth G. 1998. "Complex Instruction." *European Journal of Intercultural Studies* 9: 127-131.

Cohen, Elizabeth G., Rachel A. Lotan Beth A. Scarloss, and Adele R. Arellano. 1999. "Complex Instruction: Equity in Cooperative Learning Classrooms." *Theory into Practice* 38: 80-86.

Corra, Mamadi. 2005. "Separation and Exclusion: Distinctly Modern Conceptions of Power?" *Canadian Journal of Sociology* 30:41-70.

Corra, Mamadi and David Willer. 2002. "The Gatekeeper." *Sociological Theory* 20:180-207.

Dunn, Cynthia M. and Gary Chadwick. 2001. *Protecting Study Volunteers in Research: A Manual for Investigative Sites*. Boston: CenterWatch.

Einstein, Albert. [1934] 1954. "On the Method of Theoretical Physics." Pp. 270-276 in Carl Seelig (ed.) *Ideas and Opinions*. New York: Crown.

Fermat, Pierre. [1662] 1896. "Letter to Cureau de la Chambre." Pp. 457-463 in Paul Tannery and Charles Henry (eds.), *Oeuvres de Fermat*, vol. 2. Paris: Gauthier-Villars et Fils.

Fisher, Ronald A. 1935. *The Design of Experiments*. London: Oliver and Boyd.

——. 1956. *Statistical Methods and Scientific Inference*. Edinburgh: Oliver and Boyd.

Flament, Claude. 1962. *Applications of Graph Theory to Group Structure*. Englewood Cliffs, NJ: Prentice-Hall.

Freese, Lee and Jane Sell. 1980. "Constructing Axiomatic Theories in Sociology." Pp. 263-368 in Lee Freese (ed.), *Theoretical Methods in Sociology: Seven Essays*. Pittsburgh: University of Pittsburgh Press.

Galilei, Galileo. [1636] 1954. *Dialogues Concerning Two New Sciences*. Henry Crew and Alfonso DeSalvio (tr.) New York: Dover.

Geller, Daniel M. 1982. "Alternatives to Deception: Why, What, and How?" Chapter 2, Pp. 39-55 in Joan E. Sieber (ed.), *The Ethics of Social Research: Surveys and Experiments*. New York: Springer-Verlag.

Haney, Craig, Curtis Banks, and Philip G. Zimbardo. 1973. "Interpersonal Dynamics in a Simulated Prison." *International Journal of Criminology and Penology* 1:69-97.

Harary, Frank, Robert Z. Norman, and Dorwin Cartwright. 1965. *Structural Models: An Introduction to the Theory of Directed Graphs.* New York: John Wiley and Sons.

Heath, Thomas L. 1897. "Archimedes." Pp. *xv-xxiii* in Thomas L. Heath (ed.), *The Works of Archimedes.* New York: Dover.

Hempel, Carl G. 1966. *Philosophy of Natural Science.* Englewood Cliffs, NJ: PrenticeHall.

Hippocrates. [400 BC] 1923. *Ancient Medicine; Airs; Waters; Places; Epidemics* 1&3; *The Oath; Precepts; Nutriment.* W. H. S. Jones (tr.). Cambridge, MA: Harvard University Press.

Holmes, David S. 1976a. "Debriefing After Psychological Experiments: I. Effectiveness of Postdeception Dehoaxing." *American Psychologist* 31:858-867.

——. 1976b. "Debriefing After Psychological Experiments: I. Effectiveness of Postdeception Pesensitizing." *American Psychologist* 31:868-875.

Hughes, Everett C. 1945. "Dilemmas and Contradictions of Status." *American Journal of Sociology* 50:353-359.

Humphreys, Laud. 1970. *Tearoom Trade: Impersonal Sex in Public Places.* Chicago: Aldine.

Jones, James H. 1981. *Bad Blood: The Tuskegee syphilis Experiment.* New York: Free Press.

Kelman, Herbert C. 1967. "Human Use of Human Subjects: The Problem of Deception in Social Psychological Experiments." *Psychological Bulletin* 67:1-11.

Kruglanski, Arie. 1975. "The Human Subject in the Psychology Experiment: Fact and Artifact." Pp. 101-147 in Leonard Berkowitz (ed.), *Advances in Experimental Social Psychology*, Volume 8. New York: Academic Press.

Kuwabara, Ko. 2005. "Nothing to Fear But Fear Itself?: Fear of Fear, Fear of Greed and Gender Effects in Two-Person Asymmetric Social Dilemmas." *Social Forces* 84:1257-1272.

Leahey, Erin. 2005. "Alphas and Asterisks: The Development of Statistical Significance Testing Standards in Sociology." *Social Forces* 84:1-24.

Lederman, Leon. 1993. *The God Particle: If the Universe Is the Answer. What Is the Question?* With Dick Teresi. New York: Delta.

Ledyard, John. 1995. "Public Goods: A Survey of Experimental Research." Pp. 111-194 in John Kagel and Alvin Roth (eds), *The Handbook of Experimental Economics.* Princeton, NJ: Princeton University Press.

Lewin, Kurt, Ronald Lippitt, and Ralph K. White. 1939. "Patterns of Aggressive Behavior in Experimentally Created 'Social Climates'." *Journal of Social Psychology* 10:271-299.

Lichter, Daniel T., Diane K. McLaughlin, and David C. Ribar. 1997. "Welfare and the Rise in Female-Headed Families." *American Journal of Sociology* 103:112-143.

Lieberson, Stanley. 1985. *Making It Count.* Berkeley: University of California Press.

Lieberson, Stanley and Freda Lynn. 2002. "Barking up the Wrong Branch: Scientific Alternatives to the Current Model of Sociological Science." *Annual Review of Sociology* 28:1-19.

Lucas, Jeffery W. 2003. "Theory-Testing, Generalization and the Problem of External Validity."

Sociology Theory 21:236-253.

Lucas, Jeffrey W., Corina Graif, and Michael J. Lovaglia. 2006. "Misconduct in the Prosecution of Severe Crimes: Theory and Experimental Test." *Social Psychology Quarterly* 69:97-107.

McNeill, Paul. 1993. *The Ethics and Politics of Human Experimentation*. Cambridge, UK: Cambridge University Press.

——. 1997. "Paying People to Participate in Research: Why Not?" *Bioethics* 11:390-396.

Mann, Michael. 1986. *The Sources of Social Power*. New York: Cambridge University Press.

Mantell, David M. 1971. "The Potential for Violence in Germany." *Journal of Social Issues* 27: 101-112.

Marx, Karl. [1857] 1973. *Grundrisse*. New York: Vintage.

——. [1867] 1967. *Capital*. New York: International Publishers.

Milgram, Stanley. 1965. "Some Conditions of Obedience and Disobedience to Authority." *Human Relations* 18:57-75.

——. 1974. *Obedience to Authority*. New York: Harper & Row.

Mill, John S. [1843] 1967. *A System of Logic*. London: Longmans, Green.

Moore, James C. 1968. "Status and Influence in Small Group Interactions." *Sociometry* 31: 47-63.

Nagel, Ernest. 1961. *The Structure of Science*. NewYork: Harcourt Brace and World.

Nagel, Jack. 1975. *The Descriptive Analysis of Powey*. New Haven, CT: Yale University Press.

National Institutes of Health. 1979. *The Belmont Report: Ethical Principles and Guidelines for the Protection of Human Subjects of Research*. Washington, DC: Department of Health, Education and Welfare.

Newton, Isaac. [1686] 1966. *Principia Mathematica*. A. Motte (tr.). Berkeley: University of California Press.

Orne, Martin T. 1962. "On the Social Psychology of the Psychological Experiment: With Particular Reference to Demand Characteristics and their Implications." *American Psychologist* 17:776-783.

——. 1969. "Demand Characteristics and the Concept of Quasi-Controls." In Robert Rosenthal and Ralph L. Rosnow (eds.), *Artifact in Behavioral Research*. New York: Academic Press.

Piliavin, Irving M., Judith Rodin, and Jane A. Piliavin. 1969. "Good Samaritanism: An Underground Phenomenon." *Journal of Personality and Social Psychology* 13:289-299.

Popper, Karl R. 1962. *Conjectures and Refutation*. New York: Basic Books.

——. 1994. *The Myth of Framework*. London: Routledge.

Prechel. Harland. 2000. *Big Business and the State*. Albany, NY: State University of New York Press.

Rosenholtz, Susan J. and Elizabeth G. Cohen. 1983. "Back to Basics and the Desegregated School." *Elementary School Journal* 83:515-527.

Rosenthal, Robert. 1966. *Experimenter Effects in Behavioral Research.* New York: Appleton-Century-Crofts.

Rosenthal, Robert and Lenore Jacobson. 1968. *Pygmalion in the Classroom.* New York: Holt, Rinehart and Winston.

Schachter, Stanley and Jerome E. Singer. 1962. "Cognitive, Social, and Physiological Determinants of Emotional State." *Psychological Review* 69:379-399.

Selltiz, Claire, Marie Jahoda, Morton Deutsch, and Stuart W. Cook. 1959. *Research Methods in Social Relations*, Revised Edition. New York: Holt, Rinehart and Winston.

Sieber, Joan E. 1992. *Planning Ethically Responsible Research: A Guide for Students and Internal Review Boards.* Newbury Park, CA: Sage.

Simpson, Brent. 2003. "Sex, Fear, and Greed." *Social Forces* 82:35-52.

Skvoretz, John and David Willer. 1991. "Power in Exchange Networks: Setting and Structural Variations." *Social Psychology Quarterly* 54:224-238.

Smith, Cedric M. 2005. "Origin and Uses of *Primum Non Nocere*—Above All, Do No Harm!" *Journal of Clinical Pharmacology* 45:371-377.

Strodtbeck, Fred L., Rita M. James, and Charles Hawkins. 1957. "Social Status in Jury Deliberations." *American Sociological Review* 22:713-719.

Stouffer, Samuel A., Edward A. Suchman, Leland C. Devinney, Shirley A. Star, and Robin M. Williams, Jr. 1949. *The American Soldier: Studies in Social Psychology in World War II.* Princeton, NJ: Princeton University Press.

Thye, Shane R., David Willer, and Barry Markovsky. 2006. "From Status to Power: New Models at the Intersection of Two Theories." *Social Forces* 84:1471-1495.

Voss, David. 1999. "What Ever Happened to Cold Fusion?" *Physics World* (http://physicsweb.org/articles/world/12/3/8/1 [accessed 4 December 2003]).

Wagner, David and Joseph Berger. 1985. "Do Sociological Theories Grow?" *American Journal of Sociology* 90:697-728.

Walker, Henry A. 2002. "Three Faces of Explanation: A Strategy for Building Cumulative Knowledge." Pp. 15-31 in Jacek Szmatka, Michael Lovaglia, and Kinga Wysienska (eds.), *The Growth of Social Knowledge.* Westport, CT: Praeger.

Walker, Henry A., Shane R. Thye, Brent Simpson, Michael Lovaglia, David Willer, and Barry Markovsky. 2000. "Network Exchange Theory: Recent Developments and New Directions." *Social Psychology Quarterly* 63:324-337.

Walker, Henry A. and Morris Zelditch, Jr. 1993. "Power, Legitimation, and the Stability of Authority: A Theoretical Research Program." Pp. 364-381 in Joseph Berger and Morris Zelditch, Jr. (eds.), *Theoretical Research Programs: Studies in the Growth of Theory.* Stanford, CA: Stanford University Press.

Weber, Max. [1896] 1976. "The Social Causes of the Decay of Ancient Civilization." In

Russell Kahl(ed.),*Studies in Explanation*. R. Frank(tr.). Englewood Cliffs,NJ:Prentice-Hall.

——. [1911]1951. *The Religion of China*. New York:Free Press.

——. [1918]1968. *Economy and Society*. Berkeley:University of California Press.

——. [1904]1958. *The Protestant Ethic and the Spirit of Capitalism*. Talcott Parsons(tr.). With a foreword by R. H. Tawney. New York:Scribner.

Weber,Stephen J. and Thomas D. Cook. 1972. "Subject Effects in Laboratory Research: An Examination of Subject Roles,Demand Characteristics,and Valid Inference." *Psychological Bulletin* 77(4):273-295.

Wilkinson,Martin and Andrew Moore. 1997. "Inducement in Research." *Bioethics* 11:373-389.

Willer,David. 1967. *Scientific Sociology*. Englewood Cliffs,NJ:Prentice-Hall.

——. 1984. "Analysis and Composition as Theoretic Procedures." *The Journal of Mathematical Sociology* 10:241-270.

——. 1987. *Theory and the Experimental Investigation of Social Structures*. New York:Gordon and Breach.

——. 1999. *Network Exchange Theory*. Westport,CT:Praeger.

——. 2003. "Power-at-a-Distance." *Social Forces* 81:1295-1334.

Willer,David and Bo Anderson(eds.). 1981. *Networks Exchange and Coercion:The Elementary Theory and its Applications*. New York:Elsevier.

Willer,David,Brent Simpson,Jacek Szmatka,and Joanna Mazur. 1997. "Social Theory and Historical Explanation." *Humboldt Journal of Social Relations* 22:63-84.

Willer,David and Murray Webster. Jr. 1970. "Theoretical Concepts and Observables." *American Sociological Review* 35:748-757.

Zelditch,Morris,Jr. 1992. "Interpersonal Power." Pp. 994-1001 in Edgar F. Borgatta and Marie L. Borgatta(eds.),*Encyclopedia of Sociology*. New York:Macmillan.

Zimbardo. Philip G. Christina Maslach,Craig Haney. 2000. "Reflections on the Stanford Prison Experiment:Genesis,Transformations,Consequences." Chapter 11,Pp. 193-237 in Thomas Blass(ed.),*Obedience to Authority: Current Perspectives on the Milgram Paradigm*. Mahwah,NJ:Erlbaum.

万卷方法总书目

万卷方法是我国第一套系统介绍社会科学研究方法的大型丛书,来自中国社科院、北京大学等研究机构和高校的两百余名学者参与了丛书的写作和翻译工作。至今已出版图书60多个品种,其中绝大多数是2007年以来出版的新书。

2010年版

1 **实验设计原理:社会科学理论验证的一种路径**
戴维·威勒　亨利·沃克 著
杜伟宇　孟琦 译
书号:978-7-5624-4947-8

2 **社会统计学**
布莱洛克 著　沈崇麟 等译
书号:978-7-5624-5253-9

2009年版

3 **结构方程模型——AMOS的操作与应用**
泰勒 等著　葛道顺 译
书号:978-7-5624-4947-8

4 **质性研究:反思与评论(第贰卷)**
陈向明 主编
书号:978-7-5624-4936-2

5 **标准化调查访问**
福勒 著　孙龙 等译　张小劲 校
书号:978-7-5624-5062-7

6 **混合方法论:定性方法与定量方法的结合**
阿巴斯·塔沙克里,查尔斯·特德莱 著
唐海华 译　张小劲 校
书号:978-7-5624-5110-5

7 **量化研究与统计分析:SPSS中文视窗版数据分析范例解析**
邱皓政 著
书号:978-7-5624-4821-1

8 **如何解读统计图表:研究报告阅读指南**
纳迪 著　汪顺玉 等译
书号:978-7-5624-4906-5

9 **泰利的街角:一项街角黑人的研究(个案阅读)**
列堡 著　李文茂 等译
书号:978-7-5624-4937-9

10 **心理学研究的要义**
埃文斯 著　苏彦捷 等译
书号:978-7-5624-5098-6

11 **客厅即工厂(个案阅读)**
熊秉纯 著　杜芳琴 译
书号:978-7-5624-4886-0

12 **校长办公室的那个人:一项民族志研究(个案阅读)**
沃尔科特 著　杨海燕 译
书号:978-7-5624-4880-8

13 **质性研究的理论视角:一种反身性的方法论**
艾尔维森 等著　陈仁仁 译
书号:978-7-5624-4889-1

14 **解释互动论(校订新译本)**
邓金 著　周勇 译　刘良华 校
书号:978-7-5624-4936-2

15 **社会评估:理论、过程与技术**
泰勒 等著　葛道顺 译
书号:978-7-5624-4975-1

16 **调查研究方法(校订新译本)**
福勒 著　孙振东 等译　沈崇麟 校
书号:978-7-5624-3289-0

17 **质性研究中的访谈:教育与社会科学研究指南**
赛德曼 著　周海涛 译
书号:978-7-5624-4679-8

18 **案例研究方法的应用(校订新译本)**
殷著　周海涛 译
书号:978-7-5624-3278-4

19 **教育研究的方法论探索**
孙振东著
书号:978-7-5624-4649-1

20 **参与观察法**
乔金森著　张小山译
书号:978-7-5624-4616-3

21 **分析社会情境:质性观察和分析方法**
洛夫兰德等著　林小英译
书号:978-7-5624-4690-3

22 **建构扎根理论:质性研究实践指南**
卡麦兹 著　边国英 译　陈向明 校
书号:978-7-5624-4747-4

23 **公共管理定量分析:方法与技术(第2版)**
袁政 编著
书号:978-7-5624-3640-9

24 **AMOS与研究方法**
荣泰生 著
书号:978-7-5624-4806-8

25 **文化研究:民族志方法与生活文化**
格雷 著　许梦云 译　高丙中校
书号:978-7-5624-4698-0

26 **质性研究方法:健康及相关专业研究指南**
普拉尼等著　郑显兰译
书号:978-7-5624-4720-7

27 **如何做质性研究**
希尔弗曼 著　卢晖临 等译
书号:978-7-5624-4697-2

2008年版

28 **社会科学研究的思维要素(第8版)**
赫文著　李涤非 译
书号:978-7-5624-4465-7

29 **应用Stata做统计分析(Version 9)**
汉密尔顿著　郭志刚译

书号:978-7-5624-4483-1

30 实用抽样方法
亨利著　沈崇麟译
书号:978-7-5624-4487-9

31 哲学史方法论十四讲
邓晓芒著
书号:978-7-5624-4446-6

32 质性研究:反思与评论
陈向明主编
书号:978-7-5624-4462-6

33 社会研究方法
仇立平著
书号:978-7-5624-4456-5

34 质性资料的分析:方法与实践(第2版)
米尔斯,休伯曼著　卢晖临译
书号:978-7-5624-4426-8

35 实用数据再分析法(第2版)
利普西著　刘军译
书号:978-7-5624-4296-7

36 质性研究的伦理
丁三东译
书号:978-7-5624-4304-9

37 叙事研究:阅读、倾听与理解
利布里奇著　王红艳译
书号:978-7-5624-4303-2

38 质化方法在教育研究中的应用(第2版)
麦瑞尔姆著　于泽元译
书号:978-7-5624-4349-0

39 爱上统计学(第2版)
萨尔金德著　史玲玲译
书号:978-7-5624-4196-0

40 复杂调查设计与分析的实用方法(第2版)
雷同能著　王天夫译
书号:978-7-5624-4290-5

41 美国心理协会写作手册(APA格式)(第5版)
美国心理协会著
书号:978-7-5624-4130-4

2007年版

42 做自然主义研究:方法指南
欧兰德森著　李涤非译
书号:978-7-5624-4259-2

43 多层次模型分析导论(第2版)
Ita kreft 著　郭志刚译
书号:978-7-5624-4060-4

44 评估:方法与技术(第7版)
罗希著　邱泽奇译
书号:978-7-5624-3994-3

45 焦点团体:应用研究实践指南(第3版)
克鲁杰著　林小英译
书号:978-7-5624-3990-5

46 质的研究的设计:一种互动的取向(第2版)
马克斯威尔著　朱光明 译　陈向明 校
书号:978-7-5624-3971-4

47 组织诊断:方法、模型和过程(第3版)
哈里森著　张小山译
书号:978-7-5624-3055-1

48 民族志:步步深入(第2版)
费特曼著　龚建华译
书号:978-7-5624-3996-7

49 分组比较的统计分析(第2版)
廖福挺著　高勇译　沈崇麟校
书号:978-7-5624-3942-4

50 抽样调查设计导论(第2版)
扎加,布莱尔著　沈崇麟译
书号:978-7-5624-3943-1

51 定性研究(第1卷):方法论基础(第2版)
邓津等主编　风笑天等译
书号:978-7-5624-3851-9

52 定性研究(第2卷):策略与艺术(第2版)
邓津等主编　风笑天等译
书号:978-7-5624-3286-9

53 定性研究(第4卷):解释、评估与描述的艺术及定性研究的未来(第2版)
邓津等主编　风笑天等译
书号:978-7-5624-3948-6

54 定性研究(第3卷):经验资料收集与分析的方法(第2版)
邓津等主编　风笑天等译
书号:978-7-5624-3944-8

55 研究设计与写作指导:定性、定量与混合研究的路径(第2版)
克雷斯威尔著　崔延强译
书号:978-7-5624-3644-7

56 社会网络分析法(第2版)
约翰·斯科特著　刘军译
书号:978-7-5624-2147-4

57 公共政策内容分析方法:理论与应用
李钢著
书号:978-7-5624-3850-2

2007年以前版

58 论教育科学:基于文化哲学的批判与建构
申仁洪著
书号:978-7-5624-3641-6

59 复杂性科学的方法论研究
黄欣荣著
书号:978-7-5624-3825-0

60 社会科学研究:方法评论
陈向明著
书号:978-7-5624-3689-4

61 电话调查方法:抽样、筛选与监控(第2版)
拉弗拉卡斯著　沈崇麟译
书号:7-5624-3441-7

62 科学决策方法:从社会科学研究到政策分析
沃恩著　沈崇麟译
书号:7-5624-3669-X

63 研究设计与社会测量导引(第6版)
米勒著　风笑天译
书号:978-7-5624-3295-1

64 量表编制:理论与应用
德维利斯 著
李红 等译
书号:7-5624-3280-5

65 案例研究:设计与方法(第3版)
殷著　周海涛译
书号:978-7-5624-3266-1